DOCUMENT RESUME

ED 127 351 95 TM 005 442

AUTHOR Brophy, Jere E.; And Others
TITLE Criterion Referenced Observational Measurement in the Classroom.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO UTR&D-R-76-5
PUB DATE [Apr 76]
CONTRACT NIE-C-74-0089; OEC-6-10-108
NOTE 111p.

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; Classroom Research; Codification; *Criterion Referenced Tests; *Data Collection; Data Processing; Elementary Secondary Education; *Interaction Process Analysis; *Measurement Goals; Measurement Techniques; Observation; Research Criteria; *Research Design; Research Needs; Short Courses; Student Teacher Relationship; Teacher Education; Training Techniques
IDENTIFIERS First Grade Reading Group Coding System; Junior High School Study Coding System; Solitary Play Coding System; Teacher Child Dyadic Interaction Coding System; Texas Teacher Effectiveness Study Coding System

ABSTRACT

Presented as a mini-course on the design of classroom observation systems, this report stresses the necessity of determining the goals of any study before designing the procedures. Only if the researcher has concretely formulated the questions to be answered, can an observation system be developed that will provide the appropriate data. The importance of a period of preobservation; the choice of an appropriate coding system; and the techniques of data reduction and tabulation are thoroughly discussed. The Solitary Play Coding System, the Teacher Child Dyadic Interaction Coding System, the Texas Teacher Effectiveness Study Coding System, the Junior High School Study Coding System, and the First Grade Reading Coding System are used to illustrate the various complexities of coding systems currently in use. (BW)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED127351

CRITERION REFERENCED OBSERVATIONAL MEASUREMENT IN THE CLASSROOM[1]

The Research and Development Center for Teacher Education

UTR&D

The University of Texas at Austin

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM005 442

CRITERION REFERENCED OBSERVATIONAL MEASUREMENT IN THE CLASSROOM[1]

The main objective of the CORRELATES OF EFFECTIVE TEACHING PROJECT is to expand the number of teaching principles based on documented findings from systematic classroom research. The problems and processes studied have been selected on the basis of observation and consultation with teachers and school personnel. Emphasis is on the study of the classroom to discover how these processes can be conducted to the greatest advantage of teachers and individual students.

One of the project's major efforts was a two-year study of teaching effectiveness involving the examination of the classroom behavior of teachers consistent in producing student learning gains.

Since 1974 three other major data collection efforts were initiated and completed.

(1) STUDENT ATTRIBUTE STUDY which looked at student characteristics and behaviors and their effects on teachers.

(2) FIRST GRADE READING GROUP STUDY, an experimental study designed to test the effectiveness of selected group management techniques in teaching reading.

(3) JUNIOR HIGH SCHOOL STUDY, a follow up on earlier work from the second and third grades of the influence of teacher characteristics and behavior on students' cognitive and affective gains.

CRITERION REFERENCED OBSERVATIONAL MEASUREMENT IN THE CLASSROOM[1]

Jere E. Brophy,

Carolyn M. Evertson,

Linda M. Anderson,

Michael C. Baum,

and

John Crawford

Research and Development Center for Teacher Education

The University of Texas at Austin

Report No. 76-5

[1]This paper is a written version of the mini training course of the same title at the annual meeting of AERA, 1976, held in San Francisco.

This project was supported by the National Institute of Education Contract OEC 6-10-108, Research and Development Center for Teacher Education, and by Contract NIE-C-74-0089, Correlates of Effective Teaching. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by that office should be inferred.

Part I:

Overview
Jere E. Brophy

1. Naturalistic observation has had a long but checkered history as a method of data collection in psychology and education. In the initial stages of inquiry about almost any topic, it is necessary for making decisions about what to measure and for generating hypotheses. Typically, investigators move to controlled experiments as soon as possible. However, we believe that much is to be gained by systematizing and formalizing naturalistic observation for use in nonexperimental situations.

2. Barker and Wright broke ground with their emphasis on the sequential and continuous recording of behavior in molar behavioral units. This "specimen record" method still has advantages and usages today, although it does not lend itself well to statistical analyses. Since then, naturalistic methods have been used in varying degrees, often dependent upon the introduction of some new method that captures interest (Bales, Flanders, OScAR & others).

3. Unfortunately, there is a tendency for methods which are invented originally as means to ends (that is, as methods for studying people) to end up as ends themselves (as objects of study).

4. Also, there is a tendency to use cheap and easy methods, especially in dissertations, because of time and financial constraints on the investigators. These are understandable motivations, but they result in the use of measurement methods for purposes for which they were not designed and may not be appropriate.

5. Instead of convenience, popularity, or some other conceptually irrelevant criterion, the criterion for selection or invention of data collection methods should be the appropriateness and usefulness of these methods for answering the questions to which an investigation is addressed. It is for this reason that we have used the term "criterion referenced observation systems" in this mini course. One of our main themes, which we will both explain and illustrate at length, is that naturalistic investigations tend to pay off most successfully when the investigators have carefully prepared observation and data collection methods to match their specific needs, rather than using someone else's observation system which may or may not be appropriate to their needs. This mini course will indicate some of the considerations involved in decision making about measurement methods, and provide demonstration and practice with several <u>different</u> data collection systems developed for answering <u>different</u> kinds of questions.

6. The mini course will concentrate on methods developed for <u>in vivo</u> coding during the course of classroom interactions by observers seated in classrooms. The information presented will have value for naturalistic studies done in settings other than classrooms and/or studies involving the coding of video tapes made in classrooms, but we will concentrate on classroom research and on the coding of classroom interaction as it occurs, because we prefer this method ourselves and are most familiar with it. We prefer it to video tapes because we believe that it is much less disruptive of normal classroom procedures, and because we have found that, with the exception of a few extremely costly and elaborate specially equipped classrooms, it is very difficult to get all of the information you need on video tapes <u>in a form in which you can code it reliably and validly</u>. However,

some problems do require video tapes, audio tapes, or some other method of preserving a permanent record of what went on. What we have to say applies to the coding of these data sources, with the additional caution that investigators should pretest their data preservation methods to see if they do in fact preserve the kind of data that they need to preserve.

7. There are many reasons for conducting naturalistic studies in classrooms and elsewhere. To us, the most important single reason is generalizability or external validity. In psychology, and especially in education, we are plagued with findings from experimental situations that do not generalize to naturalistic situations, because there are too many factors in naturalistic situations that were eliminated or controlled in the laboratory but which make a great difference in the results. Another reason, and one that we become more and more impressed with in our own work and in our reactions to the work of others, is that the field of educational research seems to need more naturalistic observation in order to develop knowledge about the relevant parameters operating in different classrooms and different situations within classrooms, to develop taxonomies for classifying variables and for moving towards a common terminology for describing the same events, and, in general, for laying the groundwork to support experimental work which is in large part premature at the moment. Although there are notable exceptions, it is probably a fair statement to say that a good portion of educational research has been conducted by borrowing paradigms from psychology, sociology, and other sciences, instead of developing them systematically through a natural progression from observational and hypothesis generating studies towards experimental studies, all within educational contexts.

In short, we seem to have skipped a necessary and important part of the typical progression of scientific knowledge about any given field, and this appears to be one major reason why so much of what is called "educational research" seems to have so little to do with education, and in particular with the needs and interests of teachers.

8. This is not to say that naturalistic research is necessarily better than experimental research, but only to say that the naturalistic "spadework" that would support the development of more successful experimental research in education has been skipped for the most part. The idealized progress of scientific knowledge in the field might go something like this:

a. Naturalistic studies, controlled and uncontrolled, develop a fund of knowledge about meaningful units in classrooms, taxonomies for identifying and methods for measuring these units, and information about intercorrelations among the units and their relationships to context variables such as grade level, subject matter, group size, and so on;

b. Increasingly refined and controlled naturalistic studies are used to develop more systematic information about the relationships among variables in the naturalistic situation, and to generate hypotheses about interrelationships and their possible causes;

c. Experimental studies then build upon correlational work by testing cause and effect hypotheses and by attempting to validate potential theoretical explanations for replicated findings;

d. As dependable findings develop, they feed into teacher education and training programs, providing the conceptual and research base for the content of these programs;

e. Research and development in teacher education and training proceeds along similar lines, so that ultimately, teaching becomes an applied science in which individuals educated about educational principles, and trained to be able to implement them as needed, do so in order to optimize the educational progress of their students. Note that, although they obviously are related closely, research on the education of young students in a variety of curriculum areas is different from, and in fact builds a foundation for, research on ways to educate and train teachers in methods of teaching. Our focus today will be on the first of these two aspects of educational research, the development of a data base relating teacher behavior to student outcomes, which can be drawn upon for the development of teacher education programs.

9. In the rest of this part of my presentation, I will make several major points about naturalistic observation in the classroom. I will illustrate them briefly at times with examples, but for the most part, I will stress conceptual aspects and a general overview of the mini course as a whole in this part of my presentation. The next four presentations will provide you with detailed information and practice in operationalizing these considerations. Finally, I will return to summarize and elaborate on some of these ideas and open the floor to questions and comments.

10. The first and most fundamental point is that the goals or objectives of research should dictate the methods used. If observation systems that exist already are ideal for answering the questions posed in a given study,

this is fine. It is even better if these methods are cheap and easy to use. However, if available methods do not enable you to collect the data you need in order to answer the questions you want to answer, you will have to develop your own data gathering system. This is relatively easy to do, if you are clear about the goals of your research and if your goals are realistic.

11. Clarity about goals is a relative matter. It will vary according to the ultimate reasons for a particular study. In any case, planning a study will be easier and more efficient to the extent that the goals have been made clear. If you want to study "good teaching," what exactly do you mean by "good teaching?" If you are interested in "affective variables," what particular affective variables did you have in mind? Did you have in mind any special type of student or educational setting? Sustained thought about such questions will help you sharpen your ideas, both about the criterion measures you want to use and about the classroom data that you want to collect and relate to these criterion measures.

12. There are many other steps and considerations involved in building criterion referenced observation systems. I will mention some of them just briefly now, in order to provide an advance organizer for some of the detailed presentations and practice opportunities that you will get in the following sessions. I will come back later to summarize them and to elaborate a bit about them. Briefly, setting up your own criterion referenced observation system involves the following:

a. Conducting "preobservations" in the classrooms in which you intend to work (or classrooms very much like them), in order to get a general idea of what goes on in these classrooms. This will tell you something about what the meaningful variables are, about the frequency with which certain varibles appear, about some of the meaningful context differences that might need to be taken into account, and about the general feasibility of your plans.

b. You may find, on the basis of such preobservation, that plans have to be changed because certain variables do not appear often enough or have some other problem that would interfere with coding them validly, that different observation systems might be needed for different classroom contexts, different schools, different grade levels, or other different situations, or that certain coding plans are not feasible because observers cannot hear and see what is going on or cannot get all of the information desired.

c. Also, you may be able to identify certain repeated and psychologically meaningful sequences of interaction that are relevant to the goals of the study and can be used as major categories for your coding system. For example, one natural interaction sequence is "teacher asks a question, student gives an answer or fails to respond, and the teacher makes some kind of feedback response or fails to do so." We believe that we achieve a much greater degree of ecological validity by keying our observations to this repeated pattern of classroom interaction, as compared to what we would achieve if we used some artificial substitute like time sampling.

13. These and other considerations involved in developing criterion referenced coding systems will be described in more detail in the following presentations by Carolyn Evertson, Linda Anderson, and Mike Baum. In each of these you will be exposed to different systems that were developed for different purposes, will hear explanations of how and why things were done the way that they were done, and will get an opportunity to briefly use the system yourself to get a "feel" for what was involved in the actual process of coding. Then, Johnny Crawford will discuss ways that you can make your research more efficient by planning your coding sheets, data tabulation methods, and data analyses methods for maximal efficiency and minimal cost. Finally, I will return to summarize the major points, make a few comments about training observers, and open the discussion to questions and comments.

Part 2:

Examples of Coding Systems of Different Complexities

Carolyn M. Evertson

This is an overview of several coding systems developed for some of our research projects. Although we will not focus on any one of them in detail, several points will be made regarding each system such as (1) describing the research questions to which they were addressed, (2) defining the major sections, and (3) pointing out how our continued experience in classroom settings led to further refinement and addition of new categories and contexts.

Examples of coding sheets used in several of the studies are included as attachments at the end of this report. We will be referring to these sheets as we talk about each of the systems, so it would be helpful to place them where they are easily accessible. We will be referring briefly to the coding sheets for each of the systems to give you an idea of the larger contexts we have dealt with, to familiarize you with the broad categories we have found recurrent in our classroom research, and to touch on the changes and adaptations we have made to accommodate changes in research focus. Notice that many of the column headings are abbreviated, and that some have unique symbols inserted. These headings were labeled this way for two reasons: economy of space and confidentiality. We have found it helpful to keep the exact variables under study confidential from onlookers, to prevent biasing the data should teachers become aware of the exact behaviors being coded.

Note that, as research questions become broader, coding systems become more complex. For example, in the first system (Solitary Play) the focus is on categorizing and recording the specific play activit.es of kindergarten children playing alone. In later systems, the research questions become broader, such as "Which of many possible teacher behaviors are related to gains in student achievement?" The possibilities for behaviors to observe were much greater here.

Solitary Play Coding System

(Attachment pages 1&2)

This system was developed to study the solitary play of young children. Most of the literature on children's play treated it as evidence of non-sociability or as an outlet for imagination and fantasy. Little had been done to study it in its own right. The first page lists the definitions of categories of solitary play which we developed to study the types of activities young children engage in when they play alone. The next page is a sample coding sheet.

One of the reasons for including this system in the discussion is to make the point that building a criterion-referenced coding system need not be an expensive operation. It can be done relatively inexpensively. The most important points are that the investigator have the goals of the study clearly in mind and preobserve to see if the variables of interest are amenable to observation.

Initially, we observed kindergarteners at play and noted the types of play they engaged in when they were not playing with a friend or a group of friends. Most of their play activities could be categorized under eight

broad types (plus an "other" category). Explanations of these categories are listed on the first sheet.

The second page has space for noting the school, time period, teacher ID and location of play area (outdoor or indoor). The children in the class are listed down the side. In the squares under "activities," the observer records the type of solitary play by number. If a child engages in that type of play for longer than five minutes, the number code is circled. The information retained, then, includes the type of activity and the time the child engaged in the activity.

Kindergarten teachers were trained to use this system in about 30 minutes, once they had familiarized themselves with it. A member of the staff coded several solitary play periods with the teachers in order to be sure that the categories were understood, coding procedures were followed, and any misunderstandings were cleared up.

The teachers themselves filled out the coding sheets. We collected them at the end of the school year. The general evidence of the study did not support the idea that solitary play is typically unhealthy, undesirable, or indicative of poor personal adjustment. The more mature forms of solitary play were by far the most frequent, while those suggesting immaturity or social withdrawal were relatively infrequent. Further, the evidence suggested that, for most children at least, solitary play is a normal and probably functionally beneficial activity rather than an indicant of poor social adjustment. (Findings from this study are reported in Moore, Evertson, & Brophy, 1974).

As we go through these and other examples of coding systems, note that the complexity increases markedly. Eventually, we will look at the Junior High School Study coding system, a system that is one of our more complex. It requires several weeks of training and practice.

First, though, we will look at two systems of intermediate complexity and difficulty. These are the Teacher-Child Dyadic Interaction System and the Texas Teacher Effectiveness Study System.

Teacher-Child Dyadic Interaction Coding System

(Attachment pages 3-4)

The first of the two coding sheets for the second system is labeled "General Class Activities." This system was developed as a result of Jere Brophy's and Tom Good's interest in the mechanisms by which expectation effects are transmitted. Rosenthal's and Jacobson's Pygmalion in the Classroom (1968) had convinced that expectation effects existed, but believed that expectations were communicated through teacher behaviors. They conducted a study using teachers' own naturalistically formed expectations to divide the classes into high, middle, and low achievement groups. Since they reasoned that the mechanisms underlying teachers' expectations must lie in observable and measurable behavior, they developed the following model as a tentative explanation for expectancy effects and as a guide to systematic research in the area.

1. Early in the year the teacher forms differential expectations for students' performance;
2. The teachers then began to treat different children in accordance with their expectations for their performance;

3. The children will respond differentially to the teachers because they are being treated differently by them, and this reactive beahvior will tend to complement and reinforce the teachers' expectations;

4. As a result, the general academic performance of some children will be enhanced while that of others will be depressed, with changes being in the direction of teachers' expectations;

5. The cumulative effects of this sequence of events will be seen in the achievement test scores at the end of the school year, which will provide objective evidence that teachers' expectations function as self-fulfilling prophecies.

This led to clarity about the goal: to see if teachers treated different students differently in ways likely to maximize the learning of high expectation students and minimize the learning of low expectation students. Among the implications of settling on this goal were the following:

a. They were interested in differential teacher behavior toward different students;

b. They wanted to know about teacher expectations for students' achievement;

c. They wanted to record teacher interactions with different students separately. In other words, the student, not the teacher or class, was to be the unit of analysis;

d. They wished to identify some variables that seemed likely candidates as mediators of teacher expectation effects.

Most classroom coding systems were keyed to the teacher and did not allow for analyses based upon individual students as the unit. At this point, they reached the stage to which this entire mini-course is addressed: their goals were clear and they knew what kind of classroom process data they

wanted to collect, but they could not find an existing coding system that would do the job. Consequently, they developed the one we are looking at here (Brophy & Good, 1970b).

In its first version, the system was designed explicitly to measure the degree to which teachers interacted differently with high versus low expectation students (on variables thought to be related to the communication of teacher expectation or self-fulfilling prophecy effects). They expected some behaviors to make a difference so they developed categories and procedures to measure them. They expected that teachers would treat students for whom they had high expectations for achievement in ways that would maximize their performance. For example, they might afford them more chances to answer questions, give them more extended feedback to incorrect answers, ask them more higher order (process) questions, praise them more for good work and behavior and criticize them less for poor work or behavior compared to students for whom they had low expectations (Brophy & Good, 1970a).

The list of variables was arrived at on the basis of common sense plus previous research and writing about student motivation and effective teaching. A major problem was the non-comparability of data from one student with data from another student. Most previous studies had ignored this by lumping data from different students into a single class mean and then using this as the score for the teacher. The few studies that had scored individual students (and even many of the studies that have been done since) used simple frequency counts of different categories of behavior or types of interactions, without any attempt to control for student differences or to take such differences into account in interpreting differential teacher behavior. So this system was developed to take into account individual students' interactions with the

teacher. The research questions dictated that the student be the unit of analysis, because of the nature of differential treatment of students. It was designed to pick up those naturally occurring events which might be the vehicles for the differential communication of teacher expectations for student performance.

Now, we will turn to the coding sheet and go across the page to show how we measured behaviors of interest. We will discuss only the major headings for now, ignoring the subcategories. Subcategories will be discussed in subsequent presentations.

Student ID number is recorded in the first line. The coding proceeds from left to right across the sheet, and each successive line represents one interaction with one student. Thus, individual student data were obtained.

There are four major divisions here which are retained in most of the systems you will see later: PUBLIC vs. PRIVATE CONTACTS; STUDENT INITIATED vs. TEACHER INITIATED; ACADEMIC vs. NON-ACADEMIC; and BEHAVIORAL CONTACTS (classroom conduct). You will see other elaborations and refinements, but for the most part, these major divisions will be seen in most systems. One main section is labeled RESPONSE OPPORTUNITIES. These are public interactions between the teacher and a single student but meant for and monitored by the whole class. They occur when the teacher asks a question demanding a verbal response from the student. The RESPONSE OPPORTUNITIES section includes provision for coding sequentially: TYPE OF QUESTION, SELECTION OF CHILD, CORRECTNESS OF ANSWER, and TEACHER FEEDBACK TO THE CHILD'S ANSWER.

Other naturally occurring events in the classroom were: PRIVATE CONTACTS, those interactions happening between teacher and child in a manner not intended for monitoring by the entire class. These contacts can be subdivided as to

whether or not the teacher initiated the contact with the child or whether the child initiated the contact first. These contacts could again be subdivided according to the nature of the business to be transacted, either academic work or procedural matters.

A third contextual category is the BEHAVIORAL CONTACTS category. These account for the teacher's attention to students' classroom behavior or misbehavior and how she deals with it. She can praise (++), warn, or criticize (--) a given students' classroom conduct.

In this system, provision also was made for READING AND RECITATION TURNS, because of the differences in the meaning of small group and general class RESPONSE OPPORTUNITIES. Since frequency of READING TURNS often was determined by how many students were in the group, students in small groups got more turns than students in large groups. The nature of the turns themselves also were different. A poor reader could have many more interactions with the teacher because of the need to correct errors, so frequency of contact would not necessarily indicate high teacher expectations for performance. Low achievers could have more interactions with the teacher than high achievers because they required more correction. Likewise, good readers were likely to have very few errors, hence fewer interactions with the teacher, during reading turns. These READING AND RECITATION turns were coded on a separate sheet in the similar manner as RESPONSE OPPORTUNITIES. Coding proceeds from left to right, and each line indicates a separate interaction, just as for the previous coding sheet. Note however, that the rest of the coding sheet is a duplicate of the first sheet. During coding of reading groups or reading turns, the coder merely moved to READING AND RECITATION SHEET and coded with it. Essentially the same information is retained on this sheet as on the

GENERAL CLASS ACTIVITIES sheet, but it could be separated in analysis; this was important because of the above mentioned reasons.

In the next two systems, you will see that, while two sheets are used, different information is collected.

Texas Teacher Effectiveness Study Coding System

(Attachment pages 5-6)

The third coding system expanded a great deal on the one we just discussed. Again, we will not dwell on the details of the subcategories, but will go over the major categories and point to the adjustments in the system made to accommodate changes in research focus.

This system was used to observe in the classrooms of teachers who had shown consistency in producing achievement gains in their students over a period of three to four years. The research focus was not on individual students, but on teachers (unlike the previous study). We were not able to code individual student behavior, so instead we put check marks in the M (Male) or F (Female) columns. Thus, at least, we got information on sex of students. However, student ID easily could be included if so desired, by recording student number under the M or F columns instead of check marks. This would be appropriate, for example, in studies where the focus was on fewer variables but data were needed on individual students. However, in the Texas Teacher Effectiveness Study (Brophy & Evertson, 1976; Brophy & Evertson, Note 1), our interest was in looking at every conceivable process behavior that might allow us to determine how teachers varying in effectiveness conducted their classrooms and dealt with their students.

Now, let's examine the major sections of the coding system. As you will see, we retained the PUBLIC RESPONSE OPPORTUNITIES section, although some of the subcategories were changed and others were expanded. We found, from observing in classrooms, that students initiate PUBLIC RESPONSE OPPORTUNITIES to some extent, and this was not taken into account in the earlier system. We included it here, feeling that the degrees to which teachers allowed students to initiate questions or comments might be related to their effectiveness in producing learning gains.

These sections are labeled STUDENT INITIATED QUESTIONS and STUDENT INITIATED COMMENTS. These could be coded further as to whether they were Called Out (CALL); whether they were Relevant (REL) or Irrelevant (IREL) to the topic being discussed; and how the teacher dealt with them. She could give them no feedback (0); delay (DELAY) until a more appropriate time; not accept the question or comment (NACPT); give brief or long feedback (BRIEF or LONG); or redirect the question to the class (RDRCT). There also is provision for academic praise (++) or criticism (--) and for BEHAVIORAL praise (++), criticism (--), or warning (WARN).

In the case of STUDENT INITIATED COMMENTS, the coding in the teacher feedback section differs slightly. The teacher can give no feedback (0), delay, or not accept, as is the case for STUDENT INITIATED QUESTIONS. However, for COMMENTS, she can accept (ACPT), integrated (INTEG) into the discussion, or allow a shift (SHIFT) in topic. The other categories remain the same for both QUESTIONS and COMMENTS.

On the second page, note that we expanded the OPINION section from the one column in the Teacher-Child Dyadic Interaction System. This was to provide for praise or criticism and for certain types of teacher feedback (NO

FEEDBACK, DISAGREE, ACCEPT, or INTEGRATE into the class discussion). The section on SELF QUESTIONS was subdivided into subject matter related vs. non-subject matter related, then further categorized into PREFERENCE and EXPERIENCE RELATED. In the previous system, SELF QUESTIONS were coded with ACADEMIC RESPONSE OPPORTUNITIES. Since SELF QUESTIONS required the student to make some non-academic contribution to the class, such as "show and tell," relating personal experiences, preferences, or feelings, there were no right or wrong answers. For SELF QUESTIONS, the answer category and most of the FEEDBACK CATEGORIES (terminal or sustaining) were not applicable. For the Texas Teacher Effectiveness Study, we separated SELF QUESTIONS from academic response opportunities, and attempted to get more information about the types of SELF QUESTIONS teachers asked.

The remainder of the second page is devoted exclusively to coding PRIVATE CONTACTS initiated by the student (CHILD-CREATED CONTACTS) or by the teacher (TEACHER-AFFORDED CONTACTS). These main categories are divided into work-related and personal or procedure related. In the previous system, TEACHER-AFFORDED and CHILD-INITIATED CONTACTS were subdivided into work related and procedural. Work related contacts also were coded for PROCESS (PCSS), an explanation of steps involved in solving the student's problem; or feedback (FB), usually a brief answer or comment to the student. Procedure related contacts were coded for occurrence only (FB). Both types of contacts also could be coded for praise or criticism.

One limitation of these observation systems is that it often is difficult for the observers to hear the nature of private contacts. When they cannot hear, the best available information is the extent or length of the contact, so when we developed the Texas Teacher Effectiveness Coding System,

we recorded Brief and Long to retain this information. We knew that long contacts most frequently entail explanations on the part of the teacher, so that recording the length of contact enabled us to get at the nature of teacher feedback indirectly in situations where observers could not hear and thus code more specifically.

The section on this sheet to the extreme right of the page entitled BEHAVIOR-RELATED CONTACTS provides for coding disciplinary contacts with students. The teacher may PRAISE (++), intervene non-verbally (NVINT), warn a student to stop what he is doing (WARN), or criticize (--). Provision also is made for coding whether the teacher's attempts were unsuccessful because she did not stop the misbehavior before it went too far and thus made a timing error (TIM), or because she disciplined the wrong offending student and thus made a target error (TARG). Provision also is made for judging whether or not the teacher overreacted to the misbehavior (OVERREACT) by using overly harsh methods for the type of misbehavior. This usually is a matter of judgment on the coder's part. These distinctions were included because of our interest in Kounin's (1970) work. We wished to find out whether or not our effective or less effective teachers differed in the ways that they managed or disciplined their classes. As it turned out, they did. Effective teachers used the techniques recommended by Kounin, so that their classes generally ran more smoothly and with fewer discipline problems (Brophy & Evertson, 1976).

Junior High School Study Coding System

(Attachment pages 7-8)

The fourth and last system we will look at is one of our most complex. This study was intended as a replication of the Texas Teacher Effectiveness Study, at the secondary grades. Many relationships we expected to find at the second and third grades did not appear. For example, advance organizers, indirect teaching, and other aspects of classroom discussions were not important in the early grades, apparently because not much discussion occurs in these grades (Brophy & Evertson, 1976). We hypothesized that they would become more important with succeeding grade levels. It was partially to check such hypotheses that replication at higher grade levels was undertaken.

Some categories which proved useful in the other system were retained, such as the ACADEMIC RESPONSE OPPORTUNITIES. A few of the feedback subcategories were rearranged and changed, but essentially, the sequence of CHILD #, DIFFICULTY LEVEL OF QUESTION, selection of respondent (SELECTION), correctness of answer (ANSWER), and feedback to students' answer (SUSTAINING or TERMINAL) was retained. These designations seemed to apply to older students as well as younger ones. The minor changes in subcategories reflected the differences in teaching-learning milieu in early primary vs. junior high school grades, as determined in preobservations, as well as improvements in the basic system which resulted from earlier experience with particular coding categories and conventions. In any case, the purpose of the Junior High School Study Coding System (Brophy, Evertson, Baum, Crawford, & Edgar, Note 2) was the same as that for the Texas Teacher Effectiveness Study System--recording "any" classroom variable that might relate to teacher effectiveness in producing

learning gains. The differences in systems reflect our maturing conceptions of how to construe and measure classroom events and our observations of grade level differences in teaching-learning milieu.

Note that this first sheet of the coding system had several new additions to enable collection of information about classroom context. Since we used this system primarily in math and English classes, we were interested in the types of lessons taught and the range of subject matter covered. Again, we were interested in finding out what related to student achievement and affective gains. We believed that one important area to record was the teachers' handling of the curriculum. In previous studies, we did not concentrate heavily on curriculum, but junior high schools are more subject matter oriented. Important differences in effectiveness might lie in the teachers' management of the subject matter.

Coders checked each type of activity the teacher engaged in when teaching either of the two subject matter areas. A section called CLASSROOM FORMAT, a categorization we added after realizing the importance of teacher differences from the effectiveness study, also was included, to allow us to record the amount of time teachers spent using different teaching techniques. We also included a small space for recording peer-tutoring situations (who were tutors and who were tutees).

We combined the STUDENT-INITIATED QUESTIONS AND COMMENTS to conserve space, since most teacher feedback was appropriate to both types of student initiated interactions. Adjustments were made where the two types of student public initiation required different feedback from the teacher. While STUDENT-INITIATED COMMENTS AND QUESTIONS were not numerous in the elementary grades, where the previous system was used, we expected more of this type of

interaction in the secondary schools. OPINION questions were placed back in the ACADEMIC RESPONSE OPPORTUNITIES section, on the supposition that these types of questions were more likely to occur in secondary schools in academic class discussions. They were relatively infrequent in elementary classrooms, and unlikely to be academic in content.

The second sheet for this coding system, and the last one we will look at in this part of the mini course, provides for coding private contacts. Again, we retained the general contexts of CHILD CREATED vs. TEACHER AFFORDED, each subdivided into work related and personal or procedure related. We attempted to capture the quality of these private contacts whenever possible by using two categories AUDIBLE (AUD) and INAUDIBLE (INAUD). From our experience in classrooms, we learned that observers could hear and hence judge the quality of these private interactions some of the time. We felt that it was important to capture what information we could, so the AUDIBLE and INAUDIBLE columns were added. When the interaction was AUDIBLE, other information about the quality of the contact also was recorded.

To the right of the second page, BEHAVIOR-RELATED CONTACTS again were included, as from previous systems. However, extra provision was made for coding the types of misbehaviors and how the teachers dealt with them. We felt that misbehaviors at the junior high level were likely to be varied, and we wished to get information on types of discipline problems and how the teacher dealt with them. Part of this concern also stemmed from our interest in collecting data on individual students. Data on certain "target" students were coded by individaul ID number (in addition to student sex), for later analysis of individual differences in teacher-student relationships. Some critical behaviors (those with asterisks) were coded whether the teacher

responded to them or not. Since the students were older than the elementary age group studied earlier, many more of these behaviors occurred.

Finally, we included a column for SOCIAL CONTACTS. This section provided for recording whether the contact was child initiated (CI) or teacher initiated (TA), and whether the contact was ACCEPTED (AC) or NOT ACCEPTED (NA). The social section was included because we observed that many junior high students sought out their teachers for contacts just to get to know them. We believed that student affect and attitudes would relate to the degree to which teachers conversed with their students or to the degree to which teachers were approachable.

Again, the adjustments in the coding system were made on the basis of previous experience in the classroom and on the basis of questions, hunches, hypotheses, etc., about the effects that a different age group would have on the classroom climate.

Summary

In this presentation, we have attempted to show how systems can be adapted to changing research questions and changing classroom contexts. There are large jumps in complexity from the Solitary Play Coding System to the Teacher-Child Dyadic Interaction Coding System to the Junior High School Coding System. For example, as we observed individual private contacts with the teacher, whether child created or teacher initiated, we became aware that these could be looked at in several ways and that we could record more information. BEHAVIOR RELATED contacts also were differentiated more thoroughly as we began looking at older students.

Before leaving this topic, it is important to mention that there are a number of things that are not amenable to this particular type of classroom observation and measurement but that still appear to play an important part in the educational process. For example, because of the difficulty of hearing private interactions and hence coding them for type and quality, it also is difficult for an observer to record information about seatwork using one of these systems. Other means have to be devised for coding it, perhaps involving closer work with the teachers and seeing samples of student work with teacher comments.

Certain other variables also are difficult to record or infer from frequency counts, such as teacher warmth. Warmth often can be inferred from counts of smiles, pats, praise, etc. However, observers frequently can get much better reliability by rating teacher warmth on a five-point scale. Warmth, clarity, enthusiasm, and other "high inference" measures are examples of characteristics we have found difficult to specify in great detail. Nevertheless, they can be rated with good agreement by observers. Consequently, we typically supplement our low inference, elaborate coding of discrete classroom events with high inference ratings of broader teacher personality and style variables that are difficult to code this way but easy to rate on high inference scales (warmth, enthusiasm, organization, clarity, student orientation, etc.).

[1] Teachers in the text are referred to as "she" and students are referred to as "he."

Solitary Play by Kindergarteners Description of Activity Code

We are interested in finding out what children who choose to be alone during free play time are doing. The following category descriptions cover the most likely activities that you will find these children engaged in, but please use the "other" category to add anything that you observe that does not fit in the list we have provided. Do not code anything for children involved in any kind of social interactive play, or when a given child is in transition between activities, choosing what he will do. Some children will never play alone, others will probably do so just part of the time. Code just activities done alone.

As you enter the activity code numbers on the form, please circle the number if the child stays with the activity for more than 5 minutes.

1. Problem-solving or goal-directed: The child is concentrating on doing something and doesn't want to be bothered. Examples might be coloring, clay or other art work, building with blocks, playing with ordinary toys, musical instruments, working puzzles, collecting pretty pebbles, stringing beads. Please indicate specifically what he is working with under "Comments."

2. Role-taking: The child is taking on or pretending to be some adult role. He might play policeman or cowboy, play with a telephone, set up a housekeeping situation, mother dolls, be an airplane (pilot!), etc. Again, please note the specific activity under "Comments."

3. Large muscle play: running, climbing, swinging, etc. Please note the specific activity under "Comments."

4. Education oriented play: We mean this in the narrow sense of reading, writing and arithmetic, or school readiness activities. Examples are reading, doing substantive workbooks, Cuisenaire rods, felt board manipulations, complex mazes or mind puzzler toys, looking at aquarium or ant farm, watching a bug on the ground. Please indicate the specific activity under "Comments."

5. Onlooker: The child's interest is engaged in what some other children are doing (the types of things above), but he is not trying to get involved. Please indicate under "Comments" the activity which he is watching.

6. Daydreaming, wandering: The child is sitting or wandering around without paying attention to the activities of others. Perhaps he occasionally makes a stab at beginning something but never does get involved in it. (Do not, however, count as a wanderer someone who takes a couple of minutes to get started).

7. Sulking, pouting: The child is staying alone because he is made at someone, something, or the world.

8. Seeking out teacher: The child hangs on or constantly seeks contact with the adults available instead of engaging in activities or playing with other children.

9. Other: Perhaps a child might be anxious or nervous, or rejected by the other children and therefore playing alone. Children who are doing clean-up work. Please indicate under "Comments" what the child is doing.

"Comments" -- We'd particularly like you to note the type of activity the child is engaged in or observing in categories #1 through #5 and #9, but please make any note you feel might help us to understand the differences in the solitary play of different children.

School _____ Date _____ Indoor play _____

Teacher _____ Length of Period _____ Outdoor play _____

Please indicate the nature of solitary play activity next to the child's name. If the activity goes on for more than 5 minutes, circle the number. Cross off the list the names of any children who are absent for this play period. Note that there might be more than one type of solitary play per child.

Activity code:

1 = goal-directed | 4 = educational | 7 = sulking, pouting
2 = role-taking | 5 = onlooker | 8 = seeking teacher
3 = large muscle play | 6 = daydream, wander | 9 = other

Name	Act.	Act.	Act.	Act.	Comments
Andy					
And	(1)				coloring
Betty					
Carl					
David D	(3)				jungle gym
David W					
Fred	(6)	5			watched Mary + bug
Jorge	6	2			cowboy shooting Indians
Harriet					
Ian					
Jane	(4)				picture books
Jill					
Karen	7	8			
Louis	(8)				
Mark					
Mary	4				watching bug
Neal					
Oliver	(4)				collecting rocks
Patty					
Tom	(1)				sandbox
Vickie					

GENERAL CLASS ACTIVITIES

1. Class _____ 2. Date ___/___/___ 3. Start _____ 4. Stop _____ 5. Elapsed _____

6. Activity _____ 7. Attendance ____ 8. Observer _____ 9. Page ____ of ____

STOP TIME	RESPONSE OPPORTUNITIES																						CREATED							AFFORDED								
	CHILD				QUESTION				ANSWER				TERMINAL FEEDBACK									SUSTAINING				WORK					PROCEDURE				WORK							
	Discip	Direct	Open	Coll	Pcss	Prod	Choice	Self	+	+-	-	DK	++	Alignmt	0	Newgrneg	-	Pcss	Giv Ans	Ask Oth	Capt	Rept	Reph or Clue	New Q	++	Pcss	FB	-	?	++	FB	-	++	Pcss	FB	-	?	Pe Pr Cre	++	W	-

READING AND RECITATION

1. Class _____ 2. Date __/__/__ 3. Start _____ 4. Stop _____ 5. Elapsed _____

6. Reading Group _____ of _____ 7. Attendance _____ 8. Observer _____ Page ____ of _____

READING AND RECITATION

CHILD: Self | Work Rec. | Read Turn

TERMINAL FEEDBACK: + | Affirm Right | 0 | Negate Wrong | - | Pcss | Gives | Asks | Call Out

SUSTAINING: Rep | Reph or Clue | Neg Q

QUESTIONS

QUESTION: Pcss | Prod | Choice | Self

CHILD: Attempt | Call Out

ANSWER: + | +- | - | DK

TERMINAL FEEDBACK: ++ | Affirm | 0 | Negate | - | Pcss | Gives | Asks | Call Out

SUSTAIN: Rep | Reph or Clue | Neg Q

AFFORDED: ++ | Affirm | 0 | Negate | - | Pcss | Gives | Asks | Call Out | Rep | Reph or Clue | Neg Q

CREATED PRODEDURE: ++ FB --

CREATED WORK: ‡ Pcss FB = ?

AFFORDED WORK: ‡ Pcss FB = ?

Expectation:

Remarks:

3A

Attachment: page 4

School_____ Date___/___/___ Start_____ PAGE_____

Teacher_____ Observer_____ Stop_____ OF_____

RESPONSE OPPORTUNITIES

TIME	CHILD			QUEST			ANSWER				ACADEMIC							SUSTAINING		
											TERMINAL FEEDBACK									
	PREOL	NVOLL	CALL	PCSS	PCRSD	CHOIS	+	±	-	DKR	+ +	- -	0	PCSS	GIV ANS	ASK OTH	CALL	REPT	REPH or CUE	NEW Q

School_____ Date___/___/___ Start_____ PAGE_____

Teacher_____ Observer_____ Stop_____ OF_____

SUBJECT_____ TYPE OF ACTIVITY_____ TIME BEGUN_____

SELF

SUBJECT			NON-SUBJECT	
PREF	EXP	?	PREF	EXP

OPINION

NR	+	-	O	DISAG	ACPT	INTEG
	+	-				

CHILD-CREATED CONTACTS

WORK-RELATED							PERSONAL-RELATED						
CONT	PROC	+	-	DELAY	BRIEF	LONG	EXP	PROC	ACK	GRANT	DELAY	GRAN
		+	-									

TEACHER-AFFORDED CONTACTS

WORK-RELATED						PERS	PROCEDURE		
+	-	?	OBSV	BRIEF	LONG		FAVOR	MNGMT	THNKS
+	-								

BEHAVIOR-RELATED

+	NVINT	WARN	-	NOERR	TARG	TIM	OVRCT
+							

School _____ Date __/__/__ Start _____ Page # _____

Teacher _____ M ___ F ___ Observer _____ Stop _____ Of _____

Gen'l Class _____ Sm Grp _____ Sm Grp Mem # _____

Total Teacher Control Time _____ Absences _____

ACADEMIC RESPONSE OPPORTUNITIES

CHLD#		QUEST			SELECTION					ANSWER				SUS FDBK					TERM FDBK				
M	F	PRCS	PROD	CHOI	PREP	PRAT	NAOL	VOLL	CALL	+	-	DNKR	+	-	RSPT	EIMS	NQUS	NONAC	INTEG	PRCO	AVSS	KNSH	CAOL

SCHOOL _____ DATE: ___/___/___ START _____ PAGE ____

TEACHER _____ M___F___ OBSERVER _____ STOP _____ OF ____

GEN'L CLASS _____ SMALL # GRP. _____ MEMBER #'s _____

Behavior List

0 - Not observed
1 - Ind. Typ. Inapp.
2 - Socializing
3 - Late
4 - Disrupt*
5 - Sass/Defy*
6 - Vapp*
7 - Papp*
8 - Leave*
9 - Contraband*
10 - Bait*
11 - Sleep*
12 - _____
(other)

(* indicates a critical behavior)

CHILD CREATED CONTACTS

CHILD # M	CHILD # F	WORK: CONT	WORK: PRCD	WORK: +	WORK: -	Inaud: BRIEF	Inaud: LONG	Aud: DELAY	Aud: FDBK	Aud: PCSS	PERSONAL: GRANT	PERSONAL: DEALY	PERSONAL: NGRNT

TEACHER AFFORDED CONTACTS

CHILD # M	CHILD # F	WORK: +	WORK: -	Inaud: BRIEF	Inaud: LONG	Aud: OBSV	Aud: FDBK	Aud: PCSS	PRCD

BEHAVIOR RELATED CONTACTS

CHILD # M	CHILD # F	BEH #	NT SN	IGNOR	NVIT	MG - MT	THRT	NOFRR	FAR C	FIM	??	??	OVRCT

CHILD M	CHILD F	SOCIAL: TA	SOCIAL: CC	SOCIAL: F&B ANC	SOCIAL: F&B A

Reference Notes

1. Brophy, J. & Evertson, C. The Texas teacher effectiveness study: Classroom coding manual. (Rep. 4038) Austin, Texas: Research and Development Center for Teacher Education, The University of Texas at Austin, 1976.

2. Brophy, J., Evertson, C., Baum, M., Crawford, J., & Edgar, D. Junior high school study: Coding manual. (Rep. 4014) Austin, Texas: Research and Development Center for Teacher Education, The University of Texas at Austin, 1975.

References

Brophy, J. & Evertson, C. Learning from teaching: A developmental perspective. Boston: Allyn and Bacon, 1976.

Brophy, J. & Good, T. The Brophy-Good Dyadic Interaction System. In A. Simon and E. Boyer (Eds.). Mirrors for behavior: An anthology of observation instruments continued, 1970 supplement, Vol. A and B. Philadelphia: Research for Better Schools, Inc., 1970b.

Brophy, J., & Good, T. Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. Journal of Educational Psychology, 1970, 61, 365-374.

Kounin, J. Discipline and group management in classrooms. New York: Holt, Rinehart and Winston, 1970.

Moore, N., Evertson, C., & Brophy, J. Solitary play: Some functional reconsiderations. Developmental Psychology, 1974, 10, 830-834.

Rosenthal, R. & Jacobson, L. Pygmalion in the classroom. New York: Holt, Rinehart and Winston, 1968.

Background of the Study

From previous correlational work, especially the Texas Teacher Effectiveness Study, an instructional model was designed. It consisted of 22 principles of small group instruction which were concerned with management behaviors and were therefore "curriculum-free." That is, they did not instruct the teachers to use any particular materials, questioning styles, or approaches to teaching reading *per se*, but they did make suggestions for managing the small group. Ten first grade female teachers were given the model and were asked to use the principles of instruction in their teaching of reading during the year. There were also 10 other first grade teachers who served as a control group and were given no instruction with the model. This control group was included in the study because we wanted to measure natural implementation of these teaching principles. All of these teachers were observed approximately once a week for most of the school year with the coding system designed for the study. The purpose of observation was to gather implementation data to determine if the teaching behaviors which were included in the model were related to learning gains in students. Learning gains were assessed by achievement tests at the end of the year, with beginning readiness scores used as covariates. At the end of the school year, we then had a set of data for each teacher describing her interactions with each of her students in the reading group and residualized gain scores for those students. This enabled us to see what relationships existed between the two sets of data.

Instructional Principles

The complete instructional model is available from the Research and Development Center. However, in this presentation, I will focus on only two sets of principles in order to discuss how the coding system was constructed

Part 3:

An Elaborate Example: Coding Implementation Within Contexts in an Experimental Study

Linda Mahaffey Anderson

Several coding systems have been reviewed which demonstrate our use of certain basic distinctions in classroom observation such as public vs. private interactions, student vs. teacher initiations, and academic vs. non-academic interactions. This presentation will focus on the development of one coding system which was designed for a very specific purpose. Therefore, I will be talking in more precise terms about particular categories of observation and the decisions made in the process of designing the "First Grade Reading Group Coding System." In this presentation, I will only be talking about certain selected parts of the coding system. Manuals for the complete coding system are available from the Research and Development Center for Teacher Education.

The First Grade Reading Group Study, for which the coding system was designed, was different from previous classroom studies conducted by our project because it was experimental in nature. That is, we were trying to influence teachers to behave in a certain way and measure the effects of those teaching behaviors. Therefore, there were some very specific hypotheses to be tested in this study, and we needed a coding system that would measure the implementation of the teacher behaviors with which we were most concerned. In addition to measuring implementation, we wanted to collect data in such a way that context could be taken into account in interpreting implementation data. That is, we wanted all of our measures to specifically reflect the setting (the first grade reading group) as much as was possible.

to answer questions about the teaching behaviors suggested by the principles.

The two sets of principles are concerned with academic response opportunities. These are public interactions between the teacher and an individual child about an academic topic. For this particular study, we only recorded academic response opportunities that occurred in the reading group setting. As was discussed in the previous presentation, academic response opportunities have been identified as a naturally occurring unit in classroom interactions, and we have usually recorded tham as a sequence of events including the type of selection, the type of question, the quality of the child's answer, and any teacher feedback given to the child. We will see that this basic sequence of coding was retained in the First Grade Reading Group Coding System, but certain modifications were made within the categories of each major segment of the sequence, in order to obtain the specific measures in which we were interested. The overall framework suggested by earlier systems was useful to our needs, but none of our previous systems and no systems designed by other researchers were completely appropriate.

The first set of principles with which we are concerned dealt with the selection portion of the sequence (i.e., how the teacher picked a child to answer). The recommendations made to the teachers regarding selection were that they should:

1) Choose children in order of seating around the circle.

2) Minimize calling on volunteers, except for questions about personal experiences or opinions.

3) Discourage call outs at all times.

The second set of principles dealt with the last part of the response opportunity sequence: teacher feedback to children's answers. Specifically, we were concerned with the type of feedback given to incorrect answers or failures to respond. The treatment teachers were encouraged to respond to

incorrect answers in these ways:

1) Either give the answer to the child or give a clue to the child so that he or she could improve that answer. The specific technique that should be used in this situation was dependent on the type of question being asked and the pace of the lesson. Given the general principle, the teacher decided which option to use.

2) The teacher should never redirect a question to another child after one child had answered incorrectly or had failed to respond.

Designing the Coding System

Given these principles and the hypotheses concerning them, the problems facing us at the beginning of the study with regard to designing a coding system were:

1. How to measure implementation of these particular teaching behaviors (the types of selection and types of feedback to errors).

2. How to measure other aspects of the teaching situation so that the results related to implementation could be considered in context. That is, we wanted not only sheer frequency counts of the teaching behaviors (e.g., how many times did the teacher call on volunteers?), but we wanted to look at these frequencies for different situations (such as different types of questions, different lesson formats, different sizes and levels of groups) to see how general were the relationships of these behaviors to learning.

In order to answer these questions, we chose to retain the basic selection-question-answer-feedback sequence that was used in previous coding systems, but we modified the categories and also the coding procedures used in order to meet our needs of measuring implementation and context (Brophy, Mahaffey, Greenhalgh, Ogden, & Selig, Note 1).

The specific categories which will be described are defined in the handout under "Selected Categories from the First Grade Reading Group Coding System." These categories correspond to the columns on the coding sheet which have been checked in blue. The discussion will focus on these categories as if the coding system only included them. Information about all other categories and complete coding procedures are available in the manual. Note that the coding sheet is divided into two parts: a section for describing response opportunities and a smaller section headed "Behavior Contacts." These were corrections delivered for some kind of misbehavior. The coding procedure was to follow the teacher, noting all response opportunities when they occurred and certain behavioral contacts. This latter section was much less refined than the response opportunities section, because it was of less importance in this study. However, as will be shown, it was used to collect information which could be related to the teacher's management of response opportunities.

Measuring Types of Selection

Previous coding systems had included few alternatives for measuring types of selection. Generally, these only distinguished between volunteering, being called on without having volunteered, and call outs. Obviously, the first modification of these categories which was necessary for this study was to add a category to indicate ordered selection. This is labeled as ORDER on the coding sheet. This category was used whenever the teacher chose a child to answer a question because of his or her seating position. Even though this particular behavior was our primary concern, we also wanted to measure what other methods of selection were used when ordered turns were not. Therefore we included categories to measure alternatives: volunteering (VOL) and call outs (CALL).

Since we were also interested in what the teacher did to discourage call outs, we included procedures in the "Behavior Contacts" part of the coding system to record this information. Whenever a call out occurred and was accepted for its academic content but was also corrected by the teacher, this was noted. We also made provisions for noting ignoring of a call out by comparing the frequencies of call outs accepted but corrected, accepted but not corrected, and those ignored; thus, we would have information about a teacher's consistency in dealing with call outs and could see how such behavior affected their frequency. Call outs had not been examined so closely in previous coding systems, because they had not been the focus of a research question.

Measuring Types of Teacher Feedback

Since the principles about teacher feedback were concerned with giving the answer, providing a clue, or asking another child following an incorrect answer, we required implementation data as to the frequency of these teacher behaviors. The categories used to measure these were very similar to those used in previous systems, although the definitions were somewhat refined. The three alternatives included in the coding system were asking another child (AO), giving the answer (GA), and giving a clue (CLUE). Whenever a clueing type of feedback was offered, it resulted in the child's interaction with the teacher being sustained, since the teacher was essentially asking him or her another question, although a simplified one. This meant that every time a clue was given and recorded for a child the next line of coding was necessarily also for that child and his or her second attempt at answering the question.

Modifications of Coding Procedures Used in Previous Studies

Even though we were using a format similar to earlier coding systems (one line of coding equals one selection-question-answer-feedback sequence with

clueing feedback involving at least two lines) and similar categories for noting the types of feedback, we had a problem which had not been encountered before. This led to our making a major change in the structure of the coding process while retaining the previous coding format and categories.

The problem facing us was that we wanted to retain sequential information about the immediate effects of the CLUE category of feedback. That is, when a teacher used clueing feedback and therefore continued a child's interaction about a question in order to improve the answer, we wanted to know not only that this was done (the frequency of clues) but also how effective it was (did the clue actually lead to an improved answer?). We wanted this information because we were interested in the teacher behavior of clueing and its relation to learning gains. However, we realized that the relationship might not depend on the sheer frequency of use of clueing, but might also depend on how well it was done. This question had never been asked before, but since we were trying to closely examine the effects of different types of feedback to errors, we had to consider it.

When examining a coding sheet after a lesson, this information is obvious. That is, since each line of coding reflected a selection-question-answer-feedback sequence, at least two lines of coding were required to describe an interaction which involved clueing feedback. When looked at on the coding sheet, it is easy to see that feedback given as a clue which is noted on one line (as feedback) resulted in either a correct or incorrect response coded as the answer on the next line. Therefore, if the information was to be summed or tallied by looking at the coding sheets, it would be relatively simple to consider these two lines together and add up the number of times a "clue" on one line was followed by a certain type of answer on the next line. However, as will be discussed in the last presentation, we were planning to use computer tallying

to organize the data. This would be accomplished by converting each line of coding into a punched card which would be used for data tallying. It would, of course, be possible to keep all of the cards in the same sequence as the coding sheet and to look at relationships between succeeding cards, but the data processing involved in keeping track of such a sequence would present quite an effort. This is because all cards (representing all interactions) for a child were eventually grouped together to get a sum of each type of behavior over the whole year. Therefore, the sequence of interactions was destroyed when the cards were sorted for each individual student. The problem, then, was how to carry out this data tallying procedure (which was desirable) and also retain some sequencing information (which was also desirable). Somehow, we had to get as much information about sequence onto each line of coding, and therefore onto each single card, as was possible. However, we also wanted to retain the basic coding format of one line equalling a question-answer-feedback sequence, because this made coding a simple, and therefore more accurate, procedure, since it corresponded to a naturally occurring unit.

The solution to this problem occurred to us only after much trial and error and discussion about our future data analysis needs. This solution was to include categories on each line which reflected some information about the next line. We added a section following the categories of teacher feedback which was labeled on the coding sheet as Results. It included only two categories which represented whether or not a clue type of feedback recorded on that line resulted in an improved or unimproved response as recorded on the next line. This solved our problem of knowing the effect of each clue by simply looking at the line of coding which showed that a clue was used.

We had a related problem which necessitated a new approach to coding selections. Use of a clue type of feedback resulted in another line of coding which

would include a type of selection, question, answer, and any feedback. We therefore wanted some category of selection which would indicate that a line of coding was not an initial selection (i.e., the first or only interaction in a sequence) but was instead the result of clueing feedback attempting to improve a child's answer. In previous coding systems, a line of coding which reflected this type of continued interaction was begun by using a category of selection which also could have been used to record some initial selection. However, since one of the purposes of this study was also to examine the effects of different types of initial selection, we didn't want to use this procedure as it had been used in the past. We wanted to keep our selection categories "clean," so that codes for particular categories meant only that the child had been chosen to begin a new interaction. Therefore, we added a category to the selection section of the coding system which was used whenever an interaction, and therefore a line of coding, was the result of a teacher clue to improve a child's answer. This category was called error correction selection (ERROR) and is one of those categories checked in blue on your copy of the coding sheet. The information provided by this category is redundant when an entire coding sheet is available, but this information is not redundant when the lines have been transferred to separate cards without extensive analysis of the sequence of those cards. That is, by looking at a line preceding a second line of coding, one can determine if the second line is an initial selection or is the result of teacher effort to improve the answer. However, with the inclusion of this category of selection, the information is contained on one line.

In summary, we retained sequencing information which was important to us by these modifications of earlier methods of recording response opportunities.

By including two sets of categories on each line (the results of clueing feedback, and selection resulting from that clueing), we retained information about the effectiveness of one prescribed feedback technique when it was used, and we also could keep track of the selection variables more clearly. These two procedures were important to us simply because the behaviors of feedback and selection were important to us.

The decisions to make these changes were the result of considering future data analysis needs in the process of designing the system before data was ever collected. This consideration is important because you must know whether or not the data you are collecting with a coding system will answer the questions which you want to ask in the analysis, and because you want to be able to answer these questions in a relatively simple way. That is, the information must be retrievable from the coding in the most cost-effective way. Data analysis procedures will be further discussed by John Crawford.

Contextual Considerations in the Design of the System

Examination of the coding sheet reveals that much more information was recorded besides the categories already discussed for selection and feedback behaviors. Some of these categories measure implementation of other principles which have not been presented to you, but other categories are not directly related to any of the 22 principles. These categories were developed and used not because of a direct interest or a direct hypothesis (that is, they did not develop directly from the instructional model), but they were included because they provided specific information about the first grade reading group setting and therefore the context of those teaching behaviors in which we were more directly interested. That is, some coding categories were included because they allowed us to analyze and interpret our implementation data with respect to the surrounding events.

Keep in mind that the overall purpose of this study was to relate the teaching behaviors recommended in the instructional model to learning gains of the children. That is, we wanted to ascertain the effectiveness of the techniques recommended in the instructional model. However, we realized that answering such a question is never a simple "black and white" matter. Therefore, we wanted to phrase our questions about effectiveness in terms such as these: Is one feedback technique especially useful for a drill lesson but not for a presentation of a new story? Should the teacher try to give clues for comprehension questions, but not necessarily for questions about word recognition?

Again, I will not present all of the categories which provided contextual information, but will present only two examples. The first example of a way in which we considered the specific context of the setting is the types of questions which could be coded. Remember that each line of coding noted the type of selection, the type of question, the type of answer, and the type of feedback. We were <u>not</u> interested in influencing the teacher's types of questions asked. However, we included categories describing questions so that we could relate the types of questions asked to the selection and feedback behaviors which were of direct interest to us. Previous coding systems had been very general in the types of questions which were recorded, usually only indicating the difference between a factual question and a question requiring an explanation of a process. The question types used in this particular coding system were very specific to beginning reading. The decisions to include particular categories were based on both preliminary classroom observation and on examination of teacher manuals which suggest questions to be asked

for each lesson. In the coding system, we included nine question types which were very specific to first grade reading, such as word attack, word recognition, and comprehension questions.

Related to this was an additional problem of noting when interactions occurred during oral reading turns. This was an important contextual consideration because oral reading is a frequent occurrence in first grade reading lessons. Therefore, we wanted to be able to look separately at interactions involving reading turns and interactions involving single questions, such as comprehension questions, flash card drills, etc. Our decision again was based on future data processing needs. We decided to code interactions which occurred during reading turns in the same way that interactions about single questions were coded, but an additional column was checked for each line which indicated that it did occúr during a reading turn. Later, when each line of coding was converted into a punched card for tallying purposes, those interactions occurring in reading turns were easily spearated from others. This technique was in contrast to a method used in an earlier coding system in which reading turns were coded in a separate section of the coding sheet from coding of single questions. In an initial version of this system, we also tried this format, but discovered that it was much more difficult to code accurately this way. Therefore, we decided to do all coding of response opportunities, whether turn or non-turn, on one section of the coding sheet, but to use a single category to indicate which interactions, and therefore which lines of coding, should be considered separately. This did complicate the data tallying and analysis somewhat, but it resulted in simplified and therefore probably more accurate coding.

The second example of contextual considerations which were built into the coding system is that of recording the general lesson context. For each page

of coding, the observer noted whether or not the observation was of a workbook session, a drill, reading from the basal readers, or some other type of lesson format. This information was entered as part of the identification field for each line of coding that was converted to a punched card. This meant that interactions which occurred in different lesson contexts could be considered separately. Again, this allowed us to examine the effectiveness of certain techniques within very specific contexts. Therefore, we could see if the recommended teaching behaviors were indeed useful in some teaching situations but less useful in others. This meant that interpretation of our data would be more precise and therefore much more meaningful to teachers who would ultimately use the information.

The decision to include these general lesson contexts was probably one of the most important in the design of the coding system. However, inclusion of these lesson contexts was almost a "fluke". They were added to the coding system at the last minute, after preliminary classroom observations and practice with the coding system had begun. The observers reported that there were perhaps some important differences that were not being picked up with the system as it existed them. After discussion and just before finalization of the system, it was decided to include categories which would differentiate between lesson contexts. This example is included to reinforce the idea that trial runs with the coding system and discussion with classroom observers is very important in the design of a meaningful coding system.

There are other variables included in the coding system which did not develop directly from the instructional model, and also do not directly measure the context of the teaching behaviors of interest. However, they were included in the coding system because classroom observers noted that they were probably

of importance, and measuring them might provide useful information for further studies. Such variables were included only if measuring them did not overload the coders. That is, they represented information which was easy to collect at the same time that the rest of the coding was done, and therefore did not hinder collection of information with which we were most concerned. Examples of such variables were the way the teacher dealt with out-of-group contacts, the number of children in the group and the order of each reading group.

On the other hand, other aspects of the first grade classroom were not coded, even though they might have been of interest, because doing so would have made coding very inefficient. Examples of such variables are student initiated comments, levels of affect, and focused observation of individual children. Such variables were not included because they were not of direct importance to this study, and also because measuring them would have meant losing accuracy in recording those variables in which we were interested. These examples are presented to show that you can measure anything you decide to measure, but you must keep in mind your ultimate purposes and decide what is the most effective course to take in order to get the most useful information. It is acceptable to measure variables in which you are interested but have no specific hypotheses, but only when inclusion of these measures does not hinder the observer's recording of other more directly useful information.

Coding the sample dialogue

The categories to be used in this practive coding and the procedures for coding are described in the handout labeled "Selected Categories from the First Grade Reading Group Coding System." Please keep in mind that these categories are the only ones to be considered in this practice, but that the entire system utilizes all of the categories. Those categories checked in blue on the coding sheet in your packet correspond to these selected ones.

A completed coding sheet using these categories for the dialogue follows this discussion. Below is a description of the coding for each line of dialogue. Remember that each line begins with the child's number. Since we have not discussed types of questions and answers, we will not actually code these. However, a line drawn across these sections for each line of coding will indicate that the type of question and type of answer would be considered in more extensive coding.

Lines of dialogue	Coding:
lines 1, 2, and to slashes on line 3.	No coding. No academic interactions or behavioral contacts have taken place, and there is no place on the coding sheet to record these events.
after slashes in line 3 to slashes in line 5.	Joseph volunteered to answer the question and was called. The only teacher feedback was acknowledgement of his answer, and can not be recorded in one of the three categories of interest. Therefore, the line does not have an entry under "feedback".
after slashes in line 5 to end of line 7.	Jane called out the answer and it was accepted by the teacher. Therefore there is a line of coding indicating a "call out" selection. No feedback is noted because nothing was done

Lines of dialogue	Coding:
	which falls into one of the three relevant categories. There is also an entry made under the "Behavioral Contacts" section because the teacher also corrected the calling out behavior. In this case, the behavior was noted as #2 and the teacher reaction was a management instead of a warning, because it was patient and calm.
Between lines 7 and 8 the context of the lesson changes from #4 (New reading), when the children were answering questions preliminary to reading new material, to #3 (Drill). This change should be noted by going to a new coding sheet and marking a new context. The context for the first set of interactions should be noted at this time if it was not determined before.	
line 8	Nothing is noted about Carol and Jeff's volunteering behavior until the teacher calls on one of them. In this case, Jeff has nothing coded for him, even though he also volunteered to answer the question.
line 9 through line 11 before slash	Carol is answering because of her volunteering. Her answer received no feedback which falls into a category of interest, and so nothing is recorded under "feedback."
line 11 after slash to line 15 before slash	Jane was selected because of her seating position. (Remember that the teacher said this in line 8, and also refer to the seating order at the top of the page.) Her answer was incorrect and was followed on line 13 by a teacher attempt to help her correct it. This is recorded on Jane's first line as a "clue" and another line of coding is started for Jane because the teacher has actually asked her another question ("Now sound that out.") This second line is begun by noting that the selection was "error" and no further feedback is given which falls into one of the three categories of interest. Therefore there is no entry on the second line of coding for "feedback."

Lines of dialogue	Coding:
line 15 after slashes to line 17 before slashes	Tammy is called on because of her seating position, and therefore "order" is noted under selection. She makes an error, and the teacher responds by giving her the correct answer. This is noted by a check on that line under "gives answer."
line 17 after slashes to end of line 19	Joseph is also chosen because of his seating position and "order" is checked. His mistake is followed by the teacher asking another child to answer the question, and so there is a check on this line under "ask other."
line 20 to slashes in line 22	Jane volunteers to answer, and so "volunteer" is checked. There is no teacher feedback recorded for this line.
line 22 after slashes to slashes on line 24	Nothing is recorded about Tammy's comment because it is not an academic response opportunity or a behavioral contact. This is an example of a type of incident missed with this coding system.
line 24 after slashes to slashes on line 26	Carol called out the answer and so "call" is noted. There is no teacher feedback noted for this line, and there were no corrections given for calling out, as occurred with Jane in line 7. Therefore, nothing is entered about this call out in the "Behavioral Contacts" section.
After slashes on line 26	This teacher instruction would indicate to the coder that the lesson context was changing to #2 (workbook activities) and a new page of coding would be started if coding were to continue.

Using the Information From the Coding Sheet

If this sample of coding could be considered representative of this teacher and these children, how can we answer the questions which we asked at the beginning of the study about selection and feedback variables?

These conclusions could be drawn from this example of coding (although it would be foolish to base conclusions on such limited sampling):

1. The teacher uses volunteers and allows call outs, but when conducting a drill usually selects the children in order. In fact, ordered selection was only used during the drill, not for new reading. During "new reading" activities, there were as many call outs as volunteer selections.

2. The teacher is inconsistent in her corrections of call outs. (Jane was corrected, but Carol was not.)

3. The teacher is equally likely to use any of the three methods of responding to errors. Jane was the only child to be given a chance to improve her response when she made an error.

4. The number of response opportunities given to each child varied greatly. Jeff was never called on, Tammy had oneresponse opportunity, Carol and Joseph each had two, and Jane received four.

Given similar data on other teachers, the relationships between these teaching behaviors and individual student learning gains could be examined.

Teacher_____ Group_____ Time start_____ Page __ of __

School _____ Date _____ Time stop_____ Coder_____

No. in group_____ ____ of_____ CONTEXT: ____ 1.slow, no WB

____ 2.WB

____ 3.Drill

____ 4.New reading

____ 5.Rereading

Teacher out:
1._____Byd____
2._____Byd____
3._____Byd____

Child behaviors:
1.Callout unaccepted
2.Callout accepted
3.Indiv inapp
4.social
5.private disturb
6.posture
7.materials
8.contraband
9.other in-group
10.praiseworthy OUT-OF-GROUP
11.C init brief
12.C init long
13.T init brief
14.T init long

65

RESPONSE OPPORTUNITIES

SELECTION: Initial (NRDNM, PRE, NVOL, VOL, CALL); Subs (CONT, ERCR)

COMM, TURN, END

QUESTIONS: Read (REP, RdCHC, RdCEC, WdRET, WdATT); Non-read (PERS, CHOIC, PROD, COMP, INTRP)

ANSWERS: CORR, INC, DNK, NR

FEEDBACK: NoFB, EMPH, PRCS, PRAIS, CRIT, SPEC; Term (CA, AO, CALL); Sust (NEWQ, RPT, CLUE, GIVE)

RESULTS: IMP, NoIMP

BEHAVIORAL CONTACTS: CHD (M, F), BEH, PRAIS, INCOR, CMT, WARN, CRIT, NVINT, SPEC

CHD (M)	CHD (F)	NRDNM	PRE	NVOL	VOL	CALL	CONT	ERCR	COMM	TURN	END	REP	RdCHC	RdCEC	WdRET	WdATT	PERS	CHOIC	PROD	COMP	INTRP	CORR	INC	DNK	NR	NoFB	EMPH	PRCS	PRAIS	CRIT	SPEC	CA	AO	CALL	NEWQ	RPT	CLUE	GIVE	IMP	NoIMP	CHD (M)	CHD (F)	BEH	PRAIS	INCOR	CMT	WARN	CRIT	NVINT	SPEC
2					✓																																														
	4					✓																																													
Context changes																																																			
	5	✓			✓																																													
	4	✓						✓																														✓												
2	4	✓			✓																											✓																		
	4					✓																											✓																		
Context changes																																																			

Reference Notes

1. Brophy, J., Mahaffey, L., Greenhalgh, C., Ogden, J., & Selig, H. Coding system for the first grade reading group study. (Res. Rep. 75-2). Austin, Texas: Research and Development Center for Teacher Education, 1975.

Part 4:

Coding Affective and Behavioral Interactions: Some Examples

Michael C. Baum

In this section, we will be focusing our attention on one coding system which was developed to study how personality characteristics of students affect their interaction with teachers, and the beliefs and expectations their teachers hold about them. There are three major objectives of this section:

(1) Use it as an extended example of how a complicated observational system was developed. Demonstrating inevitable pitfalls and how the procedures proposed earlier can help you deal with them when they do arise.

(2) To make you familiar with this system. Not well enough to code, but well enough to evaluate its usefulness to you.

(3) To do some actual coding, in order to make the system understandable as well as to point out some ideas about observational systems.

The First System

As has been emphasized several times already, the coding system needs to flow from the questions you wish to answer. In this study, the questions involved determining what effect certain student characteristics had on the learning process. More specifically, we were interested in how teachers respond to different "types" of students, and to determine through the information we collected, on what basis teachers might form their beliefs.

The procedure was to identify groups of elementary school children who were consistently ranked high, medium, or low on certain personality characteristics such as intelligence, maturity, creativity, cooperation, etc., or for whom teachers held consistent attitudes (attitudes such as concern, attachment, rejection, etc.). (This study is described in more detail in Brophy, Evertson, Anderson, Baum, & Crawford, Note 1). For our purpose here, you need only know that once these groups of students had been identified, we wanted to observer their behavior in classrooms.

In thinking about what information we wanted to collect, several important classes of interactions seemed important. First, how did the student behave in typical classroom activities such as class discussion or small group activities? Secondly, we wanted to collect data on all of the other interactions, typically of a more private nature, which the students would have with the teacher--making as many distinctions among these private interactions as were both feasible and real. Thirdly, we wanted to collect information on peer interactions, assuming that many of the characteristics we were interested in would be exhibited in these interactions. Finally, we wanted to record as much of the affective quality of these interactions as possible. Once these objectives had been established, the next step was to develop an observational system which would capture all of this information, since there was not an existing system which had this potential.

The first system developed seemed to meet all of these needs. It was elaborate, particularly in its specification of the many classes of interactions which occurred in these classrooms--interactions ranging from routine classroom recitation to specification of many qualitatively different peer-to-peer interactions. It was developed after several hours of pre-observation

in the classrooms to see exactly what actually went on that was both observable and of some importance to answering our research questions.

The whole system broke down however, when the coding system was field tested by several of the staff. The system, although potentially capable of recording the behaviors of concern, was so complicated that reliability could not have been achieved.

Without going into the system in any detail, the major problems with it were: First, that it used a time sampling technique which focused on each individual target child for a brief period of time and then moved on to another. The idea was to observe a child for 30 seconds and then summarize what happened during that period of time. The problems with this were that several things often happened which then required the observer to make complex choices as to what to record. Also, frequently the target child was doing relatively little, while other students were involved in highly significant interactions with the teacher or with peers. To overcome this deficit, the rotating pattern of observation was stopped whenever so-called "critical incidents" occurred. This solved the problem of missing important events due to random focusing on children, but the observers then became confused about who to look at next and how long to focus on the critical incident. There was also the problem of equating observation time, which would have made data analysis extremely difficult. Finally, there was a problem with focusing on interactions other than ones the teacher was engaged in and therefore could affect her perceptions and/or interactions.

As a result, this initial system was retired and a new one developed. The new system incorporated many of the categories and ideas of the first, but did not have all of the associated complications. There is a copy of the final form of this system on page 16.

There were several major changes reflecting the learning involved in the pre-observation and testing of the first system, but the most fundamental was that the focus of the classroom observer's attention returned to teacher-student interaction. This approach was both psychologically easier and also reflected our goal of wanting to determine how the child affected the teacher. Even with this change however, the concept of coding events which did not involve the teacher was also retained, for a few low incidence but extremely important classroom events. The biggest deletion involved the majority of codes for student-to-student interactions.

At this point I would like to go over the observation system itself very briefly. (For further information on this system and its coding manual, see Brophy, King, Evertson, Baum, Crawford, Mahaffey, & Sherman, Note 2).

Let us begin with the Public Response Opportunities Section which is in the upper left corner of the sheet. This section is not as elaborate as others you have seen, since the focus of this study was less on teaching behavior than on teacher-student interaction.

As always in a system where it is important to retain individual student data, all coding is begun by entering the student's I.D. number. Next, a determination is made as to whether it is "small group." That is, teaching activities involving the active participation of the teacher with some segment of the entire class; or "whole class" in which the same activities are engaged in, but they involve the entire class.

The next section is used to record how the child obtained the response opportunity. Was he or she:

(1) called on without volunteering--NVOL;

(2) volunteer--VOL;

(3) wave volunteer--which was for very active attempts to get the teacher to call on the student;

(4) call out--where the child's called out answer was accepted by the teacher;

Next, optional coding of the Teacher's response of praise (+) or criticism (-) of the child's answer was recorded.

The next major section of the coding sheet involves the recording of all private dyadic interactions. In the Student Attribute Study, these private interactions became a major focus of attention. An effort was made to specify as many discrete types of interactions as possible, as well as record any affective responses if they occurred.

Again, begin with the child number. Then a determination needs to be made concerning who initiated the interaction. If it was the teacher, it is coded in the top section labeled Adult Contacts; if it was the student it was coded down below in the section labeled Child Initiated Contacts.

Let us begin by looking at Teacher Initiated Contacts. After recording the number of the child sought out by the teacher, the next task is to determine what the nature of the interaction is. Is it work related, a housekeeping request, a personal request, or merely a social contact? If you can tell, either because you can hear or because the subsequent behaviors of the participants make it clear, the interaction is coded in one of these boxes. If no such determination can be made, then the ? category is checked. This at least enables the coder to record that a teacher initiated interaction did occur with a specific child.

If it was a teacher initiated Work contact, then the next recording concerns the type of interaction. The interaction could have merely been the teacher observing a child's work without a comment OBV, or she could engage the child in either a Brief or Long interaction.

The next two sections, labeled teacher reaction and child reaction are optional codes. They are used to indicate the presence of such reactions if--and only if--they actually occur.

The teacher can respond with either praise or criticism of the child's work, and if she does they are recorded here. The child may also respond with a positive affective state indicated on the coding sheet by a happy face ☺ or with a negative affective state indicated by a sad face ☹ . Both of these are optional codes, and as such are not coded every time there is an interaction, but only when the specific event occurs.

Dropping down on the coding page--the first child initiated category is also for Work related contacts. The first thing coded reflects what happens to the request. It can be refused or the teacher can respond with brief or long feedback.

An additional section is left for any teacher reaction to the interaction. These are optional codes of praise, criticism, and a new category called impatience. This is coded when the teacher is impatient, tries to give the child the "brush-off," or in some other way indicates that she does not want to be taking the time to deal with the child's problem.

The next category of interaction in the child initiated section is also related to work, but is used when the child initiates a contact with the teacher to show-off his or her completed work. We called this Approval Seeking, since the intent of these interactions is to get some positive

feedback from the teacher. These contacts can be either accepted (FB) or refused. And the teacher may provide the child with praise or criticism of his work.

Returning to teacher initiated section at the top of the coding sheet, the next category of private interactions is called Housekeeping. These involve performing favors or errands for the teacher or tasks which must be done in the interest of carrying on the daily business of the classroom--such as passing out workbooks, cleaning the sink, etc.

A check was placed in the first column (Routine) if such a contact was made. In addition, optional coding was retained for when the teacher thanked the student for doing the task or for when the task was presented as a reward. An example of reward would be: "Johnny, you may pass the workbooks out since you got the highest grade on the test yesterday."

A child can also initiate a request to do a housekeeping task. If he/she does, it is coded down in the child initiated section. The teacher can either refuse or accept the request, with the same optional coding of reward and thanks.

The next section concerns Personal contacts. If the teacher initiates a contact with a child which involves requests to accomplish tasks which apply to that specific child alone, they are recorded here. These are requests to perform a favor, clean up a desk, etc. Routine, is coded each time the teacher makes such a contact. And once again, optional coding of any resulting affective state in the child is coded.

If the child initiates a personal request, for example asking to go to the office, the teacher's response of refusal or approval is recorded as well as a possible teacher affective reaction. For example a child may request

that he be allowed to use the learning center, and the teacher responds with a warm smile and indicates that she would very much like him to. Finally, reward is coded if the granting of the request is set up as a reward for something else.

Remaining down in the child initiated section, the next category of interactions is Tattling. When the child approaches the teacher to tattle, her response is coded, depending on whether she approves or refuses the tattle. Approval here does not necessarily mean that she acts in any way to punish the child who was told on, but merely indicates that she heard the tattle and did not respond negatively.

The last category of private dyadic interactions were called Social. This is the broadest of categories and essentially was used to record all interactions which were private, but which did not fall into any of the other categories.

Remaining in the child initiated section, social was coded for contacts which included telling the teacher about things going on at home, birthdays, new pets, etc. All that was coded here was the teacher's basic response of refusal or brief or long interaction.

However, in the teacher initiated section, optional coding of the child's affective response was also possible.

The final section we will be talking about concerns Behavior Related Contacts. This section was very much expanded because of our assumption that much of the teacher's attitudes and feelings about students would be reflected in these interactions.

So, in addition to recording the child's number and the specific responses of the teacher, many of which had been used in previous systems, we added

specific information as to the type of misbehavior engaged in as well as an optional coding of the child's affective and behavioral response to the teacher's intervention.

The specific misbehavior engaged in seemed crucial to retain since it could easily be anticipated that the teacher would have quite different feelings concerning a child who was openly defiant and aggressive than for a child whose only misbehavior involved talking to a friend.

Looking in the far right hand corner of the coding sheet, you can see a list of the specific behaviors we thought important enough to specify.

When a behavioral interaction occurred, the child's number was entered, the type of misbehavior engaged in was recorded; the teacher's response of praise, non-verbal intervention, management, warning, threat, or criticism; and then optional coding of two types of child response. The first, being cowed, was used when the child appeared sheepish, anxious, ashamed, embarrassed, afraid, remorseful, or otherwise generally sorry for having misbehaved and apparently upset at being out of the teacher's good graces. In contrast, the code sullen was used when the child appeared defiant, sullen, angry, resentful, sneering, or otherwise generally unafraid and contemptuous of the teacher.

One final thing I want to point out is that the concept of certain events being viewed as "critical incidents" was retained from the original pilot system. Thus certain behaviors were coded whenever they occurred, even if they were not seen or responded to by the teacher. These critical incidents are those circled under the student behavior list: SASS/DEFY (sassing or defying the teacher); BULL (bullying); VERB AG (verbal aggression); P AG (physical aggression); PICKED ON (recipient of bully's behavior); and DELIB TR (deliberately causing trouble).

Finally, we will not go into the last section called Adult Critical Incidents other than to say there were specific teacher behaviors which were recorded whenever they occurred such as appointing a monitor, pointing out a student as a good or bad example to the class, flattering a child, or giving physical affection.

There, very quickly you have this particular coding system. Obviously, there is much more to know before you could actually code, but this should give you some feel for what is in the system and how it works. At this point, you can do some practice coding of a few interactions, to give you more of a feel for this particular system and also to use it as a vehicle to make some points about observational systems.

Dialogue for the Student Attribute
Coding System

1st Interaction

Ms. Harris walks over to Jane's (#4) desk and asks her about the seatwork she is doing.
Ms. Harris: "The first one is not exactly right. Why don't you try adding again?"
Jane: "Oh, its 27 not 37!"
Ms. Harris: "Right! The others look fine. You're doing very well."
Jane is very obviously pleased and goes back to her work.

2nd

Joe (#2) and Jeff (#1) are in the back of the room talking about the All-star game. Ms. Harris walks over to their part of the room and stands in the aisle between them. They go back to doing their seatwork without a response.

3rd

Tammy (#3) has her hand up. Ms. Harris walks over to her desk.
Tammy: "May I clean the board Ms. Harris?"
Ms. Harris: "Well, since its obvious that you've completed the assignment before anyone else in the class, of course you can!"
Tammy smiles at Ms. Harris and says: "Thank you!"

4th

Carol (#5) begins hiccupping and in the quiet classroom the sound is very loud. After several hiccups, the class begins to snicker and giggle after each one.
Ms. Harris: "Quiet down class! You're supposed to be working in your math workbooks. Carol, please go out in the hall and get a drink (with obvious irritation in her voice) and don't come back in here until you've finished."
Carol leaves the classroom, giggling and hiccupping.

5th

Joe (#2) raises his hand and Ms. Harris goes over to his desk.
Joe: "Could you help me on this one?"
Ms. Harris:(very exasperated) "No, I want you to try it first. I really get tired of your constant requests for help, especially when you don't even try! Now get back to work and don't bother me again until you are through!"

6th

Ms. Harris walks over to Jane's (#4) desk and talks with her for a minute or two. Jane is obviously pleased when Ms. Harris leaves, but you are unable to hear their conversation because of all the noise in the room.

7th

Ms. Harris is busy working at her desk. Joe (#2) looks up, sees that Ms. Harris is busy and kicks Carol (#5) in the leg saying: "Don't ever tell Ms. Harris again when I'm copying Jeff's homework or you'll get a lot worse!"
Carol starts crying quietly at her seat.
The whole incident takes place without Ms. Harris seeing anything.

8th

Ms. Harris is approached by Jane (#4) who proceeds to tell her about her new baby brother. Ms. Harris spends a few minutes talking with Jane about this topic.

9th Carol (#5) and Joe (#2) are talking to each other in the back of the room.
Ms. Harris: "Carol, get back in your seat please. Joe, I've told you for the last time to quit goofing off. The next time I catch you up doing something other than your work, I'm sending you to the principal!"
Carol returns to her seat quickly, while Joe makes a point of taking his time, the whole time staring defiantly at Ms. Harris.

10th Jeff (#1) raises his hand and Ms. Harris comes over to his desk. He asks for help working one of the problems and Ms. Harris spends about 10 minutes outlining the procedures involved in long division and helping Jeff work a couple of problems.

11th The bell rings for the beginning of recess. Ms. Harris lines the kids up and according to her rules, all are silent except Carol (#5) who is talking to Tammy (#3). Ms. Harris says nothing, but Jane (#4) says to Tammy, "Shut up or we'll just wait here until recess is over!"
Carol stops, and the class goes out the door.

Coding Responses

1st Interaction:

* Child Number 4

It is teacher intiated
A work contact
Given brief feedback
With teacher praise
And a positive child reaction @

2nd Interaction:

Child Numbers 2 and 1

Behavior related contact
#2 social chatting
Teacher responds with non-verbal intervention
No child reaction

Note: This interaction could easily have been missed by any other than an in vivo obse-vational system or have been uncodeable if only verbal interactions were recorded.

3rd Interaction:

Child Number 3

Child initiated contact
Housekeeping request
Approved by teacher
Serving as a reward for completed work

4th Interaction:

Child Number 5

Teacher initiated
Routine personal request
No child reaction

Note: Notice that because the focus of this study was on the dyadic relationships there is no place to code the teacher's disciplining of the whole class. If this was of interest, as it was for us in the Junior High Study, a place for coding it would be needed.

* Underlined items indicate where a ✓ or number should occur on the coding sheet.

5th Interaction:

Child Number 2

Child initiated contact
Work related
Refused by the teacher
Accompanied by teacher Impatience

Note: Here is a case where the coding manual would need to be quite clear in order for coders to reliably code the teacher's affective response as impatience and not criticism.

6th Interaction:

Child Number 1

Teacher initiated
You do not know the nature of the contact ?

Note: This is a frequent occurrence in a classroom, unless the coder were to follow the teacher around, which would be quite disruptive. However, at least the fact that a teacher initiated contact with Jane can be retained. If recording exactly what happened in such contacts was important to you, then some method such as having the teacher carry a microphone would need to be used to allow access to the content of this type of interaction.

7th Interaction:

Child Number 2

Critical incident behavior number 10: "Physical aggression"
No teacher or child response

Child Number 5

Critical incident behavior number 11: being "picked on"
No teacher or child response

Note: These interactions are coded, even though the teacher was not involved. They were specified as "critical incidents" since it was our belief that some low incidence behaviors such as these could be very important to the teacher's impression or attitude if she saw them (and probably she has at other times when we were not there to record), or would be reflected in other interactions with her, or would be evident in the way other children responded to the student in interactions the teacher would see.

8th Interaction:

Child Number 4

Child initiated
Social contact
That goes on for a Long time

9th Interaction:

Child Number 5

Behavior related contact
Behavior number 2: "social chat"
Teacher management response
No child reaction

Child Number 2

Behavior related contact
Behavior number 2: "social chat"
Teacher warning response
Child defiance coded sullen

10th Interaction:

Child Number 1

Child initiated
Work contact
Given Long feedback
No teacher reaction

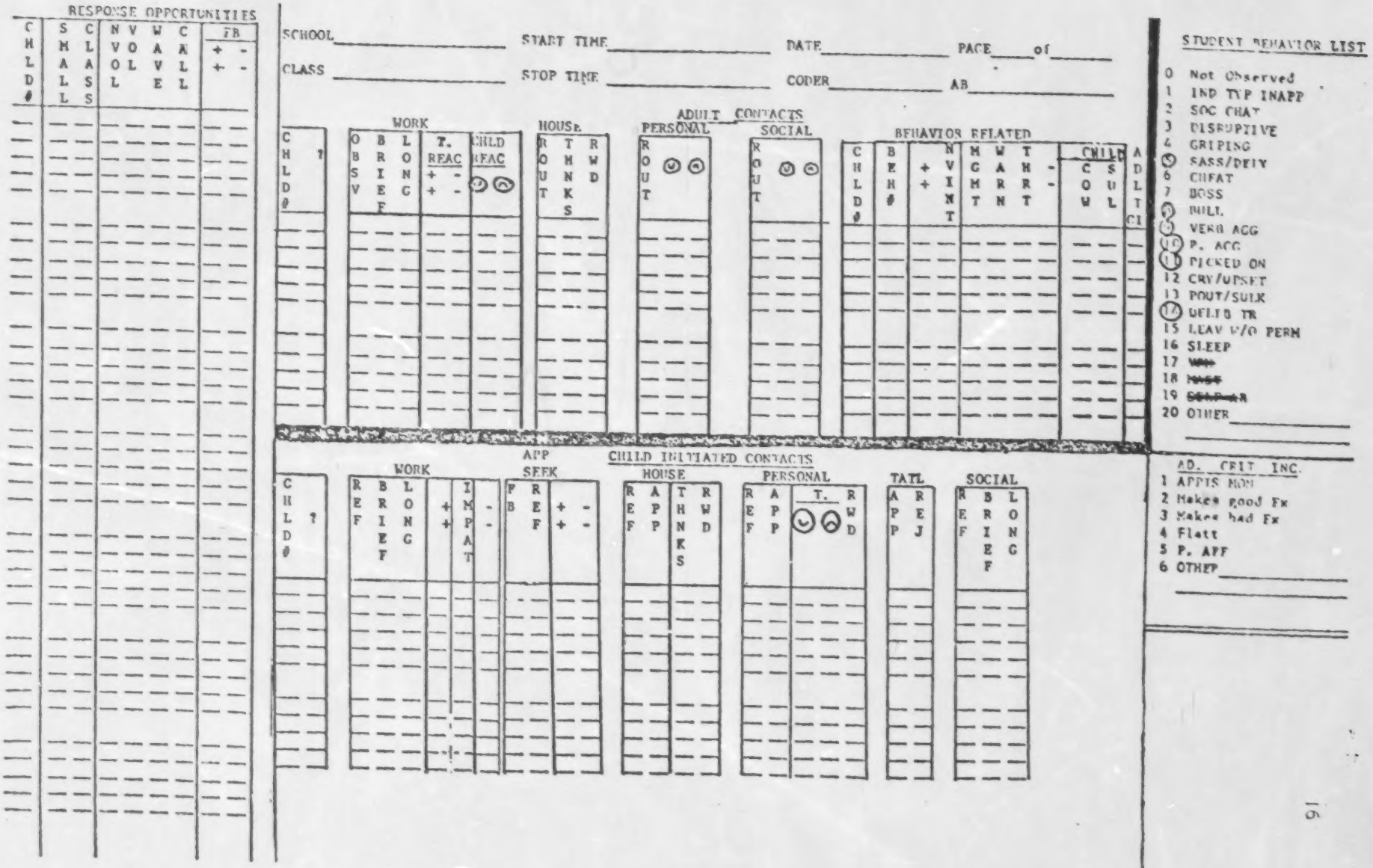
Note: Here is a case where a more detailed coding of the nature of the interaction could have been coded if the coding system was set up that way. This was not done because the more detailed information about what went on in the private work interaction was considered unimportant, but rather it was not included because it was so rarely possible for coders to make such a determination since they could not hear.

11th Interaction:

Nothing is codeable in this system.

Note: This is a rich interaction, but this coding system had no category for this behavior so it was "lost" in the low inference coding. However, coders have always been trained to make note of such interactions on blank places on the coding sheets. This enables us to either devise ways of coding these interactions by modifying the system or to at least get a better "feel" for what happens in classrooms so that if we subsequently want to develop another system, we can include categories for such interactions if it is seen as important.

(SAMPLE CODING SHEET)



Reference Notes

1. Brophy, J., Evertson, C., Anderson, L., Baum, M., & Crawford, J. The student attribute study: Preliminary report (abbreviated version). (Res. Rep. 76-1). Austin, Texas: Research and Development Center for Teacher Education, 1976.

2. Brophy, J., King, C., Evertson, C., Baum, M., Crawford, J., Mahaffey, L., & Sherman, G. Manual for the student attributes coding system. (Res. Rep. 74-2). Austin, Texas: Research and Development Center for Teacher Education, 1974.

Part 5:

Data Reduction and Tabulation

John Crawford

This report presents a discussion of a "systems" approach to reducing and tallying observational data in a manner compatible with the nature of the study and the type of data which the investigators collected. The reason as to why this type of research (naturalistic, ecological studies resulting in many process variables) requires a certain amount of planning and foresight in the production of usable frequency and proportion data may be illustrated by an example of the quantity of data which would result from one of these observational studies. Data collected from 10 one-hour observations on 30 teachers could result in approximately 10,000 interactions being coded (assuming an average of one interaction per every two minutes). Furthermore, this data might be spread over as many as 900 different students -- if the teachers had classes of approximately 30 students each. The object of the present report is to suggest ways in which experimenters may avoid being completely overwhelmed by the data once they have spent a great deal of time and energy in getting it collected.

There are two basic functions which may be served by appropriately designed tallying system:

(1) To increase the ease and efficiency of generating usable data (process data which may then be correlated, F-tested, etc.).

(2) To use the computational facilities in checking for erroneous codes made by observers during the data collection process. These "impossible codes" usually result from events occurring at a rate in excess of the observers' capacity to record them, and/or through carelessness in not dropping down to the next line or moving to an adjacent section when demanded by the situation being coded.

The implementation of a data reduction system which performs these two functions is the topic of the remainder of this discussion.

Getting the Data Tallied (Summarized) Across the Whole Experiment

Unless the study was specifically designed for the purpose of studying changes in interaction patterns from one observation to the next (across time), then the investigator will likely want to know what the totals are for each process variable using data from all observations--the number which represents the (teachers', or students', or classrooms') score for that variable for the whole study.

These numbers should be the best estimates of how the subjects would "typically" rate on such a measure. This is true to the degree that the observations did representatively sample from that classroom's history of behavior patterns. Naturally, the more observations the better; some work is currently being done to attempt to determine the minimum number of observations required to yield generalizable data. This number will surely vary for different types of variables--some do not occur with a high enough frequency (particularly in the early grades) to yield any meaningful variance even with as many as 10 or 15 observations. Other variables (e.g., number of product questions, praise after correct answer) seem capable of discriminating among teachers with only 3 or 4 observational data points.

The raw data which result from the use of dyadic interaction observational systems will have the interactions recorded by check marks under appropriate categories. The task, then, is to convert each individual checkmark into a scalar "1" and encode this information in a variable array which can be related to the category on the coding sheet. These arrays may then be cumulatively

added to yield a frequency score for the process variables. There are several ways in which this may be accomplished.

First of all, the coding sheets should be physically designed in a manner which is compatible with the tallying system which is to be set up. For example, if student sex is to be distinguished, then it might be advantageous to have two columns allocated for inserting the student identification number at the initiation of the code: under eitner "M" or "F." If this procedure were followed, and the ID numbers were punched in separate "M" and "F" fields, then information would be recorded with each interaction concerning the individual student ID (e.g., for use in correlating data by individual student) as well as the sex of the student with which the teacher interacted (yeilding data which may be easily inputted into group difference analyses of variance tests).

Also, the variables which are to be tallied should be decided on considerably prior to tallying the data. The selection of variables is obviously a crucial stage in the process since the investigators would probably not want to use data from every conceivable combination of interactions which the coding system has the ability to record. The ones which we have generally focused on are the ones which make sense (i.e., are not illogical variables which would be difficult to interpret) and also ones which are of particular experimental interest because of certain hypotheses which could be tested with data from the experiment. Another important point here is that the experimenter should make decisions ahead of time concerning "impossible" codes and train coders, as well as develop error checking routines, in an effort to minimize the occurrence of this type of data. The interactions should not even be recorded on the coding sheets themselves; but since this does occur even with highly trained coders, consideration should be given to the identification of

errors so they may be corrected.

The next stage of the data reduction process generally involves re-coding the checkmarks on the coding sheets as "1's" where there are checkmarks, and "0's" otherwise; then this raw frequency information (raw frequency by individual interaction) may be transferred to a data reduction routine. The actual summing or tallying which takes place in this data reduction routine may be done by one of several methods.

If the scope of the study is small, for example, with perhaps only two or three observations on eight or 10 teachers, it is probably best to train your coders to tally the observations by hand after each observation and then add them up to get totals for each variable across all of the two or three observations which were taken. However, if the study is of a larger magnitude than this, then it would be advantageous to develop some computational procedures for tallying the data.

One method which we have used with some success along these lines has been to keypunch the interactions directly from the coding sheets onto cards -- punching ones where there are checkmarks and blanks or zeros otherwise. This procedure then allows the summing of data on the cards (along with checking for certain errors) using a (Fortran language) tallying program. The Crawford and Washington (Note 1) technical report deals with this procedure in some depth for data resulting from one specific study -- the Student Attribute Study (see Brophy, Evertson, Anderson, Baum, and Crawford, In press).

Other methods which should work equally well that avoid the large numbers of computer cards which this procedure generates would include using a teletypewriter or an interactive cathode ray tube terminal to input the data directly from the coding sheets into a data storage system in the digital com-

puter which the investigator is using. In this instance, the data would probably be transferred directly onto a permanent file or tape which could then be called from a tallying (summing) routine capable of outputting the desired final form variables. Sackett, Stephenson, and Ruppenthal (1973) have detailed other procedures for using a magnetic tape recording device for interfacing observational data with computational facilities. This method has apparently been successfully employed with observations of chimp behavior; however, it seems to be more restricted than the keypunching or teletype input since there is a definite limit to the complexity of the type of data which may be recorded by this means -- that is, this procedure might not be compatible with naturalistic coding systems resulting in hundreds of variables.

Before final frequencies are outputted from the observational system, the data should be checked with error routines (to be discussed extensively in the next section) to see if some impossible or illogical codes were made by the observers. The use of error routines allows the experimenter to test his or her data for mistakes which may require correction or deletion from the data set. If mistakes of this type are found, (and some will almost surely be) then any changes need to be made in the raw coding sheets and on the computer cards, or permanent file, or wherever the data were stored. Then data should be rerun through the error routines to make sure that all errors have been correctly changed.

When the data is in a form such that it has all passed the error routines with no further problems then the desired frequency scores may be outputted. This will probably be done through punching out a new data deck which will include summary or total frequencies for each variable across the whole study or by writing this information onto a tape or permanent file (or both).

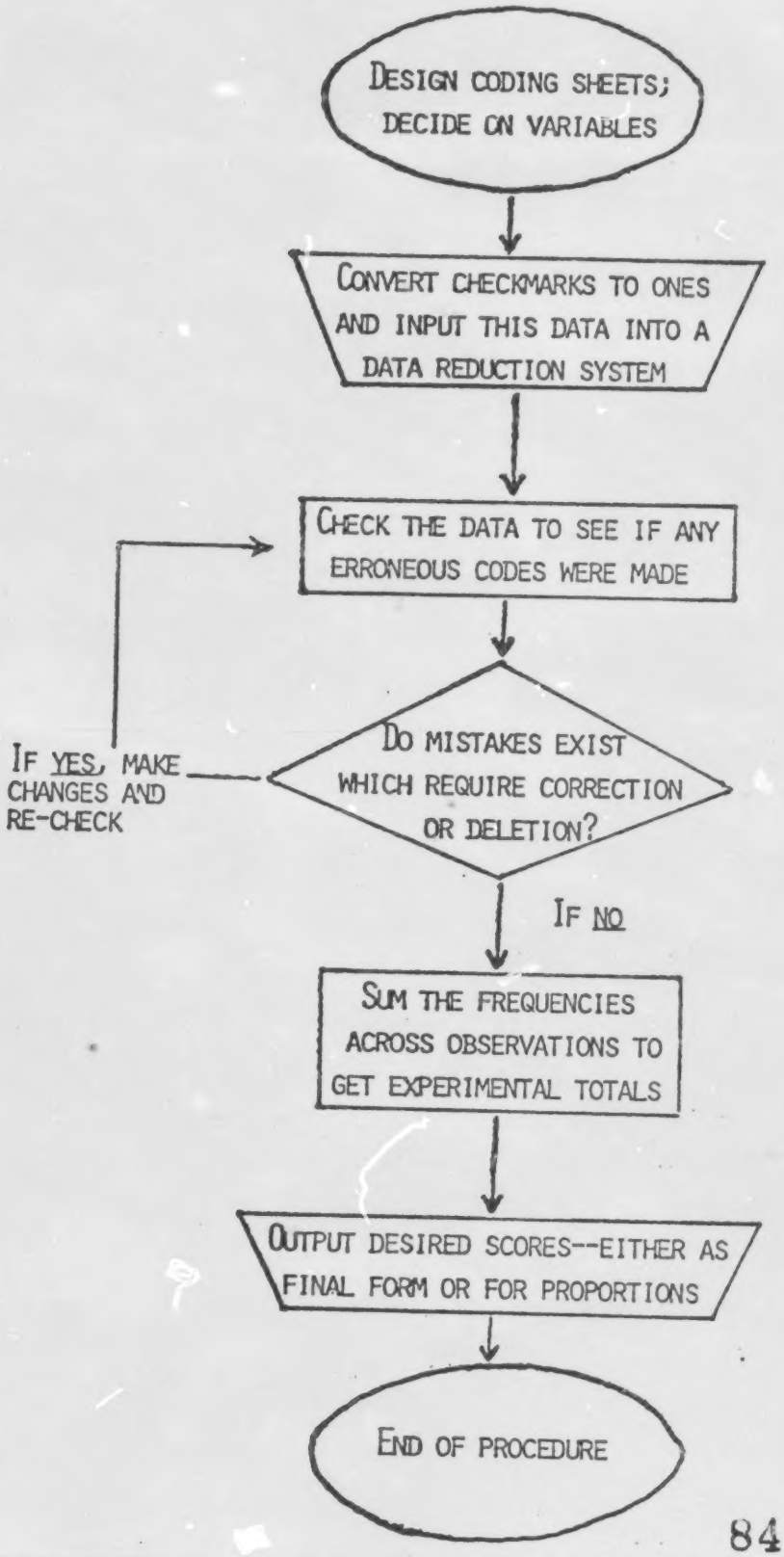
Following this, the data may be used in correlational analyses or analyses of variance tests either as final form frequency variables or for the generation of proportion-type variables.

Generally, proportion scores have been more desirable for our purposes than frequency scores (see Brophy & Good, 1970b) since teachers who vary in overall rates of interaction may be compared according to their "relative" frequencies of different types of interactions. For example, it does not make much sense to compare "number of product questions" for one group of teachers versus another if the overall rates of questioning for those two groups of teachers differed considerably. However, if the variable which is used is "proportion of all questions which are product," then valid comparisons can be made across teachers to see if there are differences in the degree to which fact-related (i.e., product) questions are asked of students in their classrooms.

Figure 1 on the next page gives a summary description of the data reduction process as proceeding from the raw coding sheets (with individual interactions noted by check marks) to final form data which may be used in statistical analyses.

FIGURE 1

SUMMARY OF DATA REDUCTION PROCESS


DESIGN CODING SHEETS; DECIDE ON VARIABLES
CONVERT CHECKMARKS TO ONES AND INPUT THIS DATA INTO A DATA REDUCTION SYSTEM
CHECK THE DATA TO SEE IF ANY ERRONEOUS CODES WERE MADE
DO MISTAKES EXIST WHICH REQUIRE CORRECTION OR DELETION?
IF YES, MAKE CHANGES AND RE-CHECK
IF NO
SUM THE FREQUENCIES ACROSS OBSERVATIONS TO GET EXPERIMENTAL TOTALS
OUTPUT DESIRED SCORES--EITHER AS FINAL FORM OR FOR PROPORTIONS
END OF PROCEDURE

Error Routines

Since the checking for errors is an important part of the system discussed here, this may be worth going into in some depth. Immediately after the data (1's and 0's) have been read by the computer program--before summing--error subroutines should be called. The first function of these routines is to locate in the array of process variables which specific fields are to be checked and to access the routines which will be used in checking. Once these fields have been located, the sequence of 1's and 0's may be inputted into the error routine to see if any impossible and/or illogical codes have been made by observers.

Examples of impossible codes would be: the observer checking both right and wrong answers for one interaction; coding both product and process questions; or putting checks under both the brief and long feedback categories within one interaction. Illogical codes would be codes which could conceivably happen but the experimenter had determined ahead of time that findings resulting from these variables would be difficult to interpret, and therefore, they should not be coded or tallied. A variable like "praise and criticism following a child's response" might be an example of this type of measure which would be difficult to interpret.

If errors are found by the checking subroutine, then changes must be made in the cards and coding sheets (and they should be run through the test routine again). The first thing which probably should be done is going back to the coding sheets and trying to figure out what happened; if the observer simply failed to skip down to the next line or omitted some information following the entry of an ID code then there may be a possibility that the interactions can be salvaged (the chances of getting the data corrected are considerably enhanced if the observer is still available for looking at the coding sheets).

Some errors will be of the type which are not amenable to correction and will simply have to be deleted. Other errors may be corrected according to some agreed upon and defensible conventions which were developed by the principal investigators. It is difficult to give firm guidelines as to what these might be, since it will be incumbent upon the investigators to defend their conventions for dealing with certain ambiguous situations occurring in the data. One example of such a convention would be the decision to use an ambiguous interaction which was coded as a process and product question as a product question. This decision could be defended on the grounds that the coding of a product question requires fewer inferences about the teacher's behavior; that is, public response opportunities are more likely to (at least) be on the "product question" level. This procedure would allow the categories pertaining to "higher order" teacher behaviors to remain uncluttered and free from any ambiguously coded interactions. The more general principle is that "conservative" conventions should be used, such that whenever an error is corrected according to an arbitrary decision-rule, the correction should be capable of being described as making fewer inferences about what the teacher (or student) was doing at the time the interaction was coded. Other coding and tallying conventions are discussed in the Brophy and Evertson (Note 2) report of the second and third grade dyadic interaction coding system.

It is perhaps worth noting here that the error routines may operate not only in "flagging" the error by print statements such as "ERROR IN X(2) AND X (3)" but also may indicate where on the coding sheets the error occurred, and during which observational period. After all data have been run through the error routines, checked, and cleaned up, the raw frequencies may be returned to the summing subroutines for final tallying. At this point, the function of the

error checking subroutines will be completed. Figure 2 is a flow chart representing the general procedures of the error checking routines.

The sample coding sheets and sample output in Figure 3 and Table 1 show observational data and outcomes of a tallying system which was used in connection with the Student Attribute Study (see Brophy, et al., Note 2, Crawford & Washington, Note 1). The "echo print" of data which preceeds the tallies of the 158 variables is useful in seeing the raw data which went into the tallying system, and also for easy checking of errors *if* an error routine had to be activated. The identification information which was recorded in this study included child number, school number, teacher number, card *type* number (corresponding to major contexts), whether the observation occurred in the morning or afternoon, and the day of the observation. The type of ID information which is recorded will vary from one study to the next according to what is necessary (what must be retrieved) for categorization of data. Also, it should be noted that it is desirable to do more than simply print out the final tallies as shown in the examples here. The best procedure is probably to punch out a new data deck *and* write the information onto a tape or permanent file system for future use in statistical analyses.

As outlined in its general form, this "systems" approach to data reduction and tallying has been tested and found to work as a means for saving considerable project time and money, and additionally in helping to find errors which may exist in some of the data sets. Hopefully, the efficient use of computational facilities and advanced input-output systems will allow this type of large-scale, naturalistic ecological study to be profitably conducted by more investigators in the future.

FIGURE 2

ERROR ROUTINE FUNCTIONS

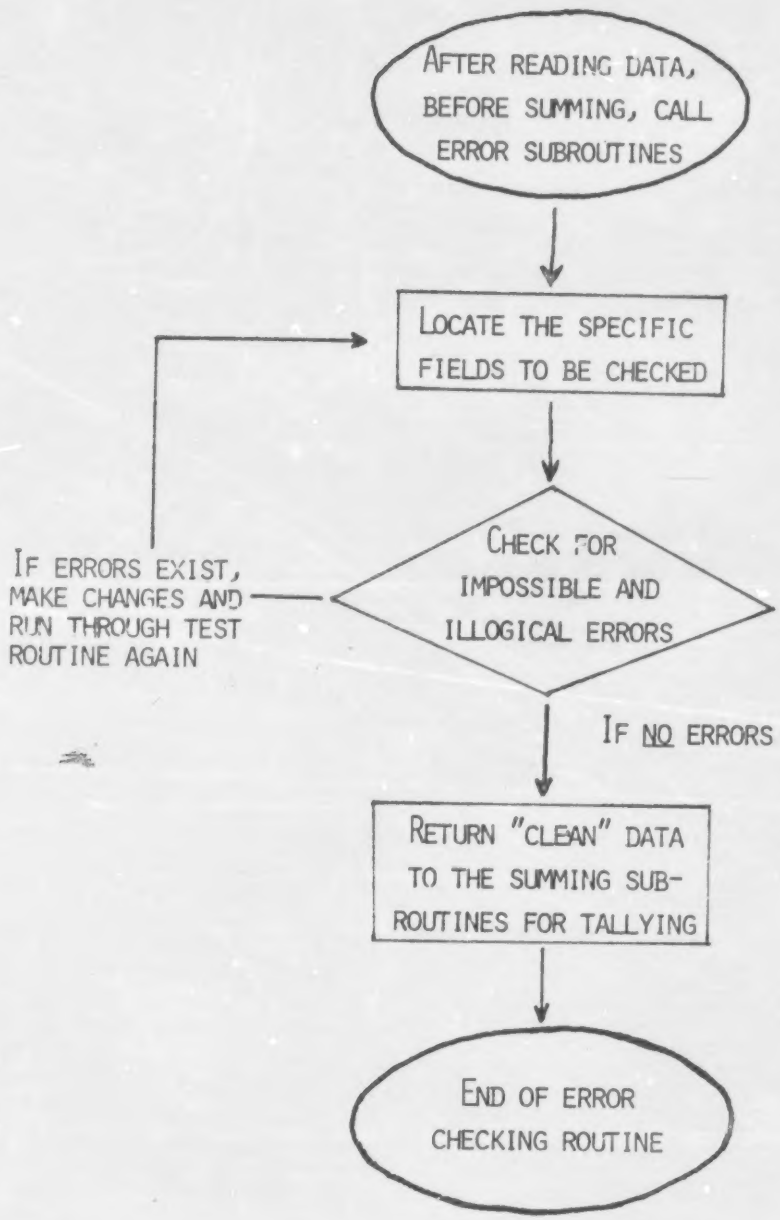
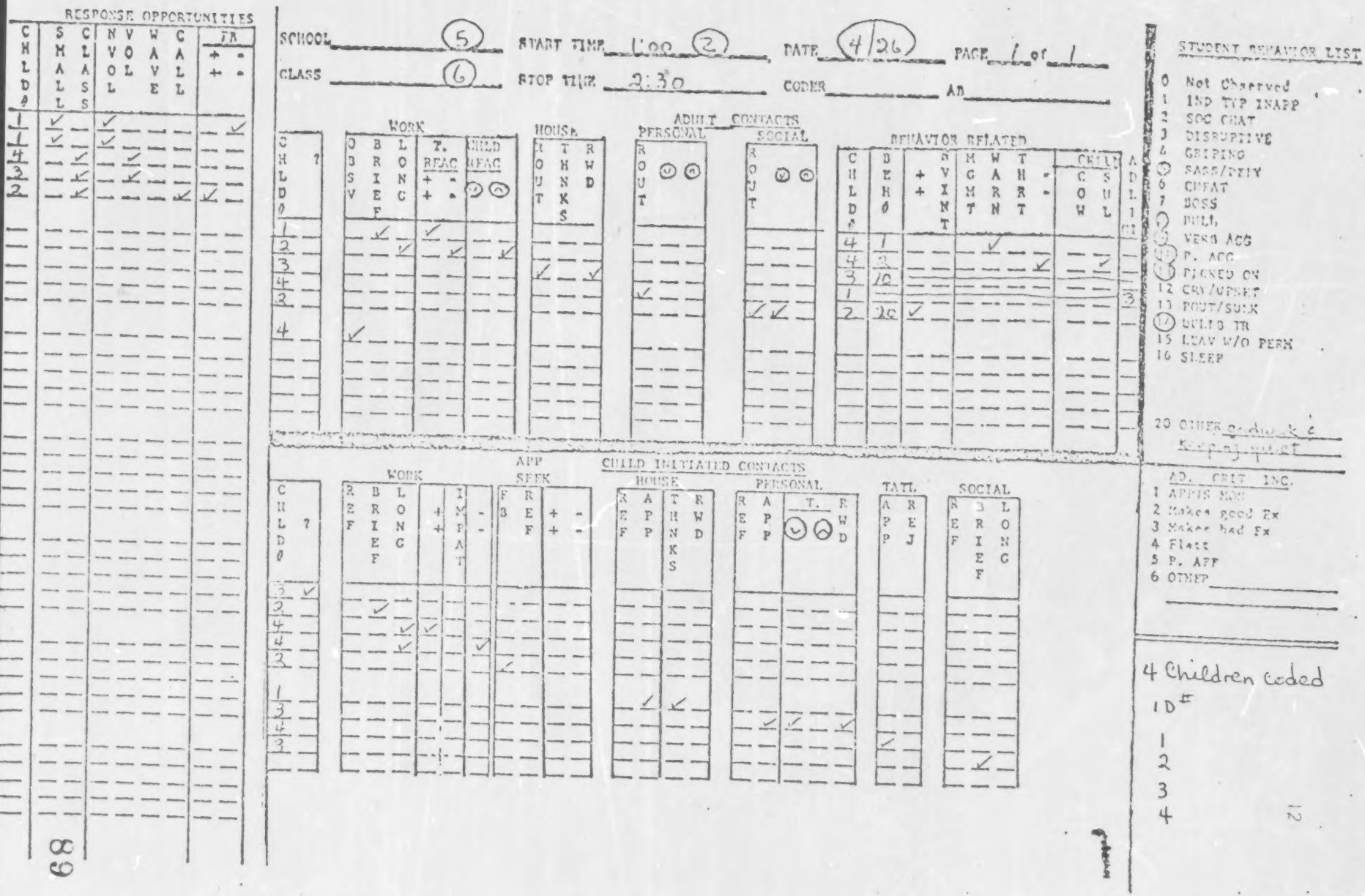

AFTER READING DATA, BEFORE SUMMING, CALL ERROR SUBROUTINES
LOCATE THE SPECIFIC FIELDS TO BE CHECKED
CHECK FOR IMPOSSIBLE AND ILLOGICAL ERRORS
IF ERRORS EXIST, MAKE CHANGES AND RUN THROUGH TEST ROUTINE AGAIN
IF NO ERRORS
RETURN "CLEAN" DATA TO THE SUMMING SUB-ROUTINES FOR TALLYING
END OF ERROR CHECKING ROUTINE

Figure 3

(SAMPLE CODING SHEET)



68

Table 1

Sample Output

```
OOL 5  AM/PM, DATE 2426 CARD 1  CHILD   1   DATA 1010****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 1  CHILD   1   DATA 1010****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 2  CHILD   1   DATA 0410****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 3  CHILD   1   DATA 0000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 4  CHILD   1   DATA 0000****************   TEACHER ID  6
                    5        10        15        20        25        30        35
 5  12426 6 2 2 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0
 5  12426 6 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
 5  12426 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5  12426 6 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
 5  12426 6 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 5

OOL 5  AM/PM, DATE 2426 CARD 1  CHILD   2   DATA 0100011****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 2  CHILD   2   DATA 0001010****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 2  CHILD   2   DATA 0000000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 3  CHILD   2   DATA 2010000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 4  CHILD   2   DATA 0010000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 4  CHILD   2   DATA 0000001000000000000000   TEACHER ID  6
                    5        10        15        20        25        30        35
 5  22426 6 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 2 1 0 0 0 0 0 0
 5  22426 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0
 5  22426 6 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5  22426 6 0 2 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
 5  22426 6 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 6

OOL 5  AM/PM, DATE 2426 CARD 1  CHILD   3   DATA 0101000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 2  CHILD   3   DATA 0000000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 3  CHILD   3   DATA 1000000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 4  CHILD   3   DATA 1000000000000000000000   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 4  CHILD   3   DATA 0000000000000000000000   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 4  CHILD   3   DATA 0000000000000000000010   TEACHER ID  6
                    5        10        15        20        25        30        35
 5  32426 6 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
 5  32426 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0
 5  32426 6 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5  32426 6 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5  32426 6 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 1 0 2 6

OOL 5  AM/PM, DATE 2426 CARD 1  CHILD   4   DATA 0101000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 2  CHILD   4   DATA 0000000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 3  CHILD   4   DATA 0100000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 3  CHILD   4   DATA 0100000****************   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 4  CHILD   4   DATA 0000000000000000000000   TEACHER ID  6
OOL 5  AM/PM, DATE 2426 CARD 4  CHILD   4   DATA 0000000000000000000000   TEACHER ID  6
                    5        10        15        20        25        30        35
 5  42426 6 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0
 5  42426 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0
 5  42426 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5  42426 6 1 3 0 2 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5  42426 6 0 0 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 8
```

Reference Notes

1. Crawford, J. & Washington, W. A computer system for tallying data collected by the student attributes coding system. (Res. Rep. 74-3). Austin, Texas: Research and Development Center for Teacher Education, 1974.

2. Brophy, J. & Evertson, C. Texas teacher effectiveness study: Classroom coding manual. (Res. Rep. 76-2). Austin, Texas: Research and Development Center for Teacher Education, 1976.

References

Brophy, J., Evertson, C., Anderson, L., Baum, M., & Crawford, J. Student personality and teaching (tentative title). Boston: Allyn and Bacon, in press.

Brophy, J. & Good, T. The Brophy-Good dyadic interaction system. In A. Simon and E. Boyer (Eds.). Mirrors for behavior: An anthology of observation instruments continued, 1970 Supplement, Volume A. Philadelphia: Research for Better Schools, Inc., 1970.

Sackett, G., Stephenson, G., & Ruppenthal, G. Digital data acquisition systems for observing behavior in laboratory and field settings. Behavioral Research Methodology and Instruments, 1973, 5, 344-348.

Part 6:

Summary and Discussion

Jere Brophy

1. If you have been clear about the goals of your research, and if this has led you to sufficiently specific and detailed identification of those criterion data and classroom information to be collected, you are ready to start the specifics of the study. The first step is to conduct preobservations in the classrooms to be included in the research (or classrooms as similar as possible to them), in order to determine the feasibility of the decisions made so far. You may find that what looked very good on paper does not work in practice. Among other things, preobservations of this kind can establish whether or not the things you want to measure happen with sufficient frequency to make efficient data collection possible, whether or not the observer can see and hear enough of the relevant interactions in order to code them accurately, and identification of relevant context differences in these particular classrooms that should be taken into account in the coding, either by eliminating certain contexts from consideration or by coding them separately.

2. The frequency with which behaviors of interest occur in naturalistic situations is relevant both to the feasibility of the study as a whole and to the question of how much observation time per class will be needed in order to get reliable information. In general, the more frequent a particular event is, the more feasible the study and the fewer observations needed per class. You may find that variables of interest just do not happen often enough to make a naturalistic study possible, even if you could visit classrooms 30 or 40 times apiece. For example, small group instruction occurs routinely in the early grades in most schools, but practically never thereafter. Consequently, if you are interested in variables relating to small group instruction,

you should study at the early grades and be aware that your findings will be restricted to these grades, at least for most practical purposes. As another example, consider the variety of questions that deal with teacher expectations and attitudes as they affect, or are affected by, individual students. If you want to study these kinds of questions, you almost have to select self contained classrooms where a single teacher deals with the same group of students for all or at least most of the day. It is possible, but very difficult, to conduct these kinds of studies in team teaching situations, both because horrendous bookkeeping problems become involved in keeping track of which interactions were with which teachers and students, and because there is no conceptually satisfactory way to allocate specific portions of ultimate student outcomes at the end of the year to the different teachers that interacted with particular students. Other aspects of feasibility were illustrated earlier, including one case where an entire coding system had to be dropped because it was not going to work. That is, it was not going to answer the questions that we wanted to answer.

These cautions about the frequency of observation of variables of interest should not be taken to mean that low frequency variables cannot be studied. They <u>can</u> be studied, and some low frequency but high intensity variables are extremely important, even though they may not happen very often. For example, a single incidence of serious violence in the classroom might have greater effects over the course of the year than hundreds of interactions involving mild affect. Again, the ultimate criterion for inclusion and exclusion of a variable is the objective of the study. If preobservation reveals that the study of a crucial variable is not feasible

according to the present plan, it will be necessary to look around for other naturalistic situations where it is feasible or, failing this, to switch to an experimental or quasiexperimental situation which will insure that the variable occurs often enough to be studied meaningfully.

3. In the course of conducting a preobservation, you will find out about whether or not the proposed coding system is feasible from the physical standpoint. That is, can the observers see and hear what they need to see and hear, and can they capture all of this reliably? If they cannot see and hear things, the coding manual will have to be revised, you will have to transfer to a different setting where they can see and hear; or, you will have to move to experiments or quasiexperiments that insure that they can see and hear. We have done a variety of things about this problem, depending upon the importance of the variable and the relative frequency of it. In the case of interactions that happen frequently but occasionally are missed because they go on completely across the room from the observers, we have simply added a "couldn't code" column, and instructed the observers to code whatever aspects of the interaction that they could code. For example, we usually distinguish between student versus teacher initiated interactions and between work and procedural interactions. Sometimes, observers will see that a student initiated interaction with the teacher, but will be unable to determine whether the interaction dealt with work or procedural matters. In this case, they can code the student for initiating an interaction with the teacher, but can code a question mark column on the coding sheet in order to indicate that they could not determine whether the interaction was a work interaction or a procedural interaction.

Where you cannot afford to miss the content of particular kinds of interactions, you will need to try to improve the situation by placing coders in better positions to see and to hear, or by moving to different classrooms or different situations. Sometimes, just moving the placement of coders will do the job. Usually, coders sit unobtrusively in the back of the room in order to code general classroom activities, and they seldom have difficulty seeing and hearing under these circumstances. However, in the early grades, when coding reading groups or other small group activities, the coders need to sit right behind the group rather than at the back of the room. This method cannot be used for coding teacher initiated private interactions with students, however, because the coder would have to physically follow the teacher around the room, something that is just too disruptive and obvious. The disruption and salience of this method probably would affect the validity of the coding of such interactions, so that we have not used it. If it were essential to hear everything that went on during such interactions, I believe that I would favor asking the teachers to wear microphones as they went around the room, and having the coder keep track of each student that the teacher interacted with in turn, so that the data would be complete when transcribed and collated later.

In any case, preobservations and trial runs of the provisional coding system will provide information about its feasibility. Most probably, it will lead to some changes in the definitions of variables included in the system and/or in the coding conventions used to take into account the context differences or define which aspects of the interaction are to be coded and which are not. These decisions should be made on the basis of the interaction of feasibility with the goals of the research. If particular variables are not crucial to the r search and also not feasible

for coding, they can be dropped or coded in a way that differs somewhat from the original plan. If feasibility prevents the coding of certain crucial variables, a fundamental shift in plans would be required in order to find a naturalistic setting or develop an experimental one that will allow the crucial variables to appear often enough to be coded and to appear in ways that allow the coders to see and hear what they need to see and hear in order to code accurately.

4. Perhaps the most important aspect of preobservations and provisional tryout of proposed coding systems is the development of information about daily routines in the classrooms of interest and about which aspects coders can see and hear and which aspects they can and cannot code reliably. It may turn out, for example, that certain daily routines were not anticipated and will require decision making as to whether or not they should be coded and how they should be coded (for example, in our teacher expectation research, we had to develop a separate method for coding reading turns, and we had to make decisions about what aspects of opening and closing rituals and routines would and would not be coded). Other important context differences include teacher versus student initiation, public versus private interactions, work versus procedural versus behavioral interactions. Some of the many other distinctions in context which have been made were illustrated in earlier presentations.

5. Preobservation, especially the information you collect about the frequency with which variables of interest are observed, will give you an idea of how long you will need to observe in each classroom in order to get reliable information. By "reliable" here, I do not mean reliability between coders;

I mean reliability in the sense of getting a sufficiently detailed sample of classroom interaction to be a reliable estimate of what goes on in that classroom. This will vary considerably, depending upon whether the unit of analysis is the teacher or the individual student and upon the frequency of the behaviors of interest. Where the focus is on teacher behavior, fewer interactions will need to be observed than when the focus is on individual students. Research using individual students as the unit of analysis is usually costly and time consuming, because you may have to observe in the classroom 20 or more hours in oruer to get reliable estimates of student status on certain variables, and you may need hundreds of hours in order to get individual student data on variables that do not occur very often. For example, if you are interested in the relative rates of praise of good behavior and criticism or punishment of misbehavior, you can get reasonably good estimates for teachers on the basis of a few observations per classroom. However, if you are interested in the rate of behavioral praise and criticism or punishment directed at individual students, you may have to observe a given classroom many times. This is especially true for behavioral praise, which occurs rarely (and in some classrooms, never).

Another outcome of preobservation should be the development of information about naturalistic classroom units. Many classroom interaction analysis methods have used time sampling, in which coders observe for a specified number of seconds and then make a recording, then observe again, then record again, and so on. This is unfortunate, because time sampling is not a psychologically valid way to unitize the flow

of spontaneous interaction. We do not interact, in classrooms or elsewhere, on the basis of time units. Instead, we interact on the basis of what Tolman used to call "molar behavioral units." Many of these can be identified by observing in the classrooms, and coding can become much more psychologically valid if the coding system is designed to follow such units. For example, one unit which is virtually universal across classrooms is "teacher asks question, student gives answer or fails to respond, and teacher reacts with some kind of feedback or response." We have included this unit in most of our classroom coding, recording each such unit that occurs by indicating the number of the student involved (if individual student data are being taken), the level of difficulty of the question asked, the method by which the student got the response opportunity, the quality of the students' response (or lack of same), and the nature of the teacher's feedback reactions. Similarly, we have found that private interactions between the teacher and individual students can be divided into interactions involving work, interactions involving procedural matters, and interactions involving teacher response to student conduct. Furthermore, some of these interactions are initiated by the teachers and some by the students. We have built these distinctions into our coding systems, also.

6. As a result, our coding sheets and manuals allow for separately coding interactions that occur in separate contexts, and for retaining the sequence of action and reaction in those interactions that involve such chains. This has proved to be important in uncovering orderly relationships among variables coded separately by context and sequence. Systems that do not make these distinctions mask these differences. As a brief example, we have found that teacher praise given in student

initiated interactions (ironically, especially in interactions initiated by students who seem to be seeking praise) tends to correlate negatively with student learning gains. In contrast, teacher praise given in interactions initiated by the teachers themselves tends to correlate positively with learning gains, if it correlates at all. Separate coding of teacher initiated versus student initiated contacts makes possible these kinds of findings.

Also, retention of the sequential nature of chains of interactions makes it possible to supplement the more typically used frequency scores with ratio or proportion scores that allow for better comparisons of individuals who differ in their rates of interactions with the teachers or of teachers who differ in their general rates of interactions with students. In addition, it makes the data much more interpretable, because one can begin to take into account student effects on teachers and student individual differences generally, something that you cannot do with pure frequency counts (except at the classroom level by sampling from very different populations). To illustrate the kinds of variables that I am discussing, and to show how they allow one to go past the level of inference which is possible from frequency measures alone, note the data on pages 17 and 18. These data are from the study mentioned earlier in which Tom Good and I tried to develop information about how teacher expectation effects might be mediated through differential treatment of different students. Note that all of the group differences on page 17 are consistent with the idea that the teachers might have been treating the high expectation students more favorably than the low expectation students. However, note also that, lacking additional information, these same data can be interpreted more simply by assuming that the high expectation students got

better teacher response for the simple reason that they were brighter and better behaved, and thus more successful in the classroom. This does not change the differential group data indicated on page 17, but it does change the interpretation of these data with reference to inferences about teacher expectation effects, teacher bias, and so on. (Brophy & Good, 1970a).

In contrast, look at the data on page 18. These are proportion scores rather than frequency scores, and they allow comparisons among students who may have differed considerably in their frequencies of particular kinds of interactions with the teachers. The basic logic here is to compute a frequency score for a variable of interest and then convert it to a proportion score by dividing it by the total number of interactions in which the variable of interest could have shown itself or could have been expected. For example, the percentage of correct answers praised is determined by dividing the number of times that a child is praised for a correct answer by the number of times that that child gave correct answers. Thus, children who are praised 10 times out of 100 possible times will get the same score as children praised 20 times out of 200 possible times. This does not allow for total comparability, because it does not take into account the probability that certain correct answers were more "praiseworthy" than other correct answers, but nevertheless, it does allow more convincing and presumably meaningful comparisons of groups than the frequency data on page 17 allow.

7. These and other points mentioned previously were elaborated in detail in the earlier presentations. I will close this presentation with some additional remarks about planning naturalistic studies, developing coding systems, and training coders.

8. First, if you are clear about your goals, you also should be clear about the kinds of classroom process information that you want to measure. This should guide you in selecting classrooms according to such variables as grade level, subject matter, self-contained versus team taught; traditional versus open space, and so on. Even so, go out and observe in the classrooms that you intend to study, both in order to make sure that the classroom events that you anticipate do in fact occur there with sufficient frequency, and to make any needed adjustments in your coding system.

9. If you are unsure about what actually goes on in the classes you want to observe in, go out and find out. Classify the interactions you observe there according to which kinds you consider appropriate for inclusion in the study (and therefore for inclusion in the coding manual), and which kinds are to be ignored (if any). Identify meaningful and repeated psychological interaction units and sequences, so that your coding system can be based on psychology meaningful variables rather than artifactual ones like time samples or frequency counts. Develop ways to build context differences into the coding system by allowing separate identification or coding of interaction occuring in different contexts, and/or by avoiding certain contexts altogether because they are irrelevant.

10. If the classrooms you had in mind do not seem very feasible for the study that you want to conduct, see if you can find classrooms that are more feasible. If you cannot, consider shifting from a naturalistic to a more controlled method of research. Or, consider shifting to another

topic, because failure to find the behavior of interest going on anywhere will mean that your topic has little external validity, at least in the schools as they are presently organized and conducted.

11. Within what is feasible, try to build into your coding system everything that is crucial to the study as defined by its basic goals. If the result is a coding system that is extremely easy to use and leaves coders with a lot of free time on their hands, you might consider building in additional variables in order to collect more information while you are at it, without adding any additional costs. On the other hand, if the data you want to collect is beyond the ability of a single coder to collect reliably, you will need to cut down the number of things that you are demanding of coders or divide your coding into two or more systems and use only one system at a time. This could be done by having more than one coder in the room at the same time, each using a different system, by having coders observe at different times using different systems, or by coding some variables with high inference ratings, depending upon feasibility and the objectives of the research.

12. Coder training should be as realistic and thorough as possible. Unless the study involves coding of video tapes, we have found that coding of video tapes is not a very useful or efficient method. The one primary advantage it has is that a single video tape can be compiled (usually this will require splicing together brief segments from a number of different tapes). This tape can be coded, and this coding then can be used as a criterion against which to compare the coding of new observers undergoing training.

However, for our purposes, we have found it most useful to train observers by:

a. Having them read the coding manual in detail and discuss it with the trainer;

b. Having them write out brief examples of each of the variables included in the system;

c. Practice coding using transcripts which we prepare specifically for practice with a particular coding system. For the most part, these transcripts give the coder practice in making the more important and obvious discriminations, although borderline cases are included as well;

d. Finally, coders are sent into classrooms, working in pairs, for practice coding. This last step usually is most successful if coders are sent into the classrooms in which they actually will be working once the study begins. This way, observers and teachers and students get acclimated to one another while the observers practice using the system, and practice is conducted in the actual situation. Decisions about when coders have reached reliability and are ready to code on their own are essentially arbitrary.

Our own method has been to use an 80% agreement criterion. Agreement percentages are computed by dividing the number of times that both observers coded an interaction and agreed by the sum of that number plus the times they both coded but disagreed, plus the number of times that one of observers coded and the other did not, plus the number of times that the second observer coded and the first did not. This is actually a stiff criterion, given the way that it is computed. Furthermore, coders are expected to reach reliability criteria for each major subpart of the coding system, not just

for the coding system as a whole. That is, they must be reliable for coding level of difficulty of questions, quality of student answers, teacher behavior during teacher initiated work interactions, and so on.

We find that most coders attain reliability within a week or two, depending upon their own aptitude for the job and the complexity of the coding system. Today, we have shown you coding systems varying from extremely simple ones to extremely complex ones, and you can see for yourself that some can be learned very easily and others require some practice. Most typically, early coding with the complex systems produces poor agreement, primarily because neither coder has built up much speed and both have difficulty keeping up with classroom events. However, this is largely a mechanical learning process, much like the initial difficulty one encounters when trying to learn to drive. After a few practice sessions, coders develop facility in finding their place on the page and keeping up with what is going on in the classroom. At this point, disagreements typically occur because coders both code the same interaction but code it differently, and not so much because one observer coded it and one did not. At this point, it makes sense to begin to compute agreement percentages and to discuss the differences. It may be that the differences are constant, occurring because one observer is coding correctly and the other is not. These cases are eliminated easily by pointing out the problem to the observers who are <u>not</u> coding correctly. However, sometimes the problem occurs because of random disagreements. Where these persist, they indicate that the coding manual is not specific enough, or that something in the nature of the coding situation makes it difficult to reliably discriminate the information called for in the manual. This

may require some adjustment in the coding manual. For example, this is how our "question mark" originated. We found that coders simply could not hear or see well enough in some cases to tell whether an interaction dealt with work or procedural matters. Rather than leave this kind of error in the coding system, we included the "question mark" column to indicate that the teacher or the student initiated an interaction, but that the coder could not tell specifically what type of interaction was involved.

Another way to handle problems like this is develop some sort of arbitrary coding convention. For example, we always code the words "Fine" and "Good" as "praise," even if they are delivered in a matter-of-fact tone. This is a purely arbitrary decision, reflecting the fact that the distinction between mere positive affirmation of a correct response and praise of that response is an arbitrary difference. We chose to set the arbitrary line where we did, and we are reasonably happy with it, although others would have reasons to put the line somewhere else. In any case, any coding system involving teacher praise necessarily involves some kind of definition of "where praise begins." That is, there must be a distinction between teacher reactions that indicate to students that they have responded correctly but which do not involve "praise," versus teacher reactions which are intense or positive enough to be defined as "praise."

As with anything else, the relative importance of these coding conventions depends upon the importance of the variables involved to the goals of the study. Where the variables are crucial, it is essential that the coding convention be successful not only in enabling coders to collect the data of interest with the coding system, but also to eliminate any problems in unreliability. On the other hand, if the behavior involved is of very low priority or even essentially irrelevant to the goals of the study, it can be eliminated from the coding system entirely or combined with other behaviors into a large category that is not given much importance in analyzing the data. For example, in some of our studies, procedural interactions in which the teachers ask the students to do classroom management tasks for them, or in which the students ask the teachers for permission to do such tasks, were considered to be essentially irrelevant to the goals of the study. In these studies, the frequency of these procedural interactions was noted, but little else about them was coded. However, in other studies, where we were interested in the frequency and quality of interactions of teachers with specific individual students, we not only coded these procedural interactions much more carefully, but we made numerous distinctions concerning the quality and nature of these interactions in addition to noting the frequency of their occurrence.

Finally, in addition to tailoring the coding system to the demands of the study, you need to give some thought to the design of coding sheets, the mechanics of data collection and tabulation, and the inputting and programming involved in data analysis. As Johnny Crawford illustrated in his presentation, seemingly minor considerations of this sort can make

a great difference in the time, difficulty, and expense of a study. In some of our own work, we have estimated conservatively that streamlining and computerizing our data tabulation and analysis routines saved us about $40,000 a year, or about 20% of our budget. Obviously, then, this is no minor matter. In addition to the financial savings, automating the data collection and tabulation saves endless hours of hand-tabulating that is boring and thankless work and that involves a high likelihood of human error, something that is eliminated with machine tabulating.

In closing, I guess the moral of the story here is "think ahead!" If you know what data you want, how you are going to get it, and what you are going to do with it when you have it, you are not only more likely to reach your goal, but more likely to reach it quickly and efficiently.

TABLE 1. Group difference in quantity and type of contacts with teachers[1]

Variable	p-values for group effects						
	Class	Sex	Expectancy	Class x Sex	Class x Expectancy	Sex x Expectancy	Class x Sex x Expectancy
Number of times child raises hand	.001	ns	.01	ns	.01	.05	.10
Number of times child is called on to answer a question	.001	ns	ns	ns	ns	.05	ns
Procedural interaction initiated by child	.001	.05	.05	.05	ns	ns	.01
Work-related interactions initiated by child	.05	ns	.01	ns	.05	ns	ns
Teacher-initiated procedural interactions	.001	ns	ns	ns	.05	ns	ns
Teacher-initiated work-related interactions	.001	.05	ns	.05	.05	ns	ns
Teacher-afforded behavioral criticisms	.05	.01	.01	ns	ns	.01	ns
Teacher-afforded questions during reading groups	ns	ns	.10	ns	.01	ns	ns
Calling out answers during reading groups	.001	ns	ns	ns	ns	ns	ns
Total teacher-afforded response opportunities	.001	.05	ns	ns	ns	ns	ns
Total child-initiated response opportunities	.001	ns	.01	ns	.05	ns	ns
Total dyadic contacts	.001	.01	ns	ns	ns	ns	ns

[1] From Brophy & Good, 1974a.

TABLE 2. Group differences on variables related to the communication of teacher expectations

	p-values for group effects						
	Class	Sex	Expectancy	Class x Sex	Class x Expectancy	Sex x Expectancy	Class x Sex x Expectancy
Number of direct questions from teacher	.01	.10	ns	ns	ns	ns	ns
Number of times called on to answer open questions/number of times child raises hand	ns	ns	.05	ns	ns	ns	ns
Percent of correct answers followed by teacher praise	.01	.01	.05	ns	ns	ns	ns
Percent of wrong answers followed by teacher criticism	.001	ns	.01	.001	.05	ns	ns
Percent of wrong answers followed by repetition of rephrasing of the question	ns	ns	.10	ns	ns	ns	ns
Percent of reading problems followed by repetition or rephrasing of the question or by giving a clue	.05	ns	.01	ns	.01	ns	ns
Percent of answers (correct or incorrect) not followed by any feedback from the teacher	.10	ns	.001	ns	.05	ns	ns

*There were no open questions in class 1.

From Brophy & Good, 1970a.

References

Brophy, J. & Good, T. Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. Journal of Educational Psychology, 1970, 61, 365-374.